THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Volume 23 Part 1 February 1971

PUBLISHED FOR THE EXPERIMENTAL PSYCHOLOGY SOCIETY
BY ACADEMIC PRESS LONDON AND NEW YORK

NOTES FOR AUTHORS

Only original papers are accepted for publication. All manuscripts should be *submitted in duplicate* to *Professor N. S. Sutherland, Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex, BN1 9QY, England.* One copy must be a *typewritten original* with double spacing and wide margins. Figures should be drawn to professional standards in Indian ink on tracing linen or Bristol Board; the original drawing (plus one copy) should be submitted. Allowance must be made for a reduction in size when the drawing is reproduced (the width of lines as reproduced should be between 0·25 and 0·33 mm). Explanations must be placed in the figure legend; the drawing should contain as little lettering as possible and lettering should be inserted in *blue* pencil or on a tracing paper to fit over the drawing. Tables and figure legends should be typed on separate sheets of paper. Where possible related diagrams should be grouped together to form a single figure.

Intending contributors should pay attention to the *Journal*'s conventions, and in particular to the form of headings, and sub-headings used, and references (in the Reference list titles of journals will in future be printed in full without abbreviation). The *Journal* will use the *Système Internationale* for all measurements (if terms like inches are used, metric equivalents must be quoted). The use of standard abbreviations in the text is permissible (if in doubt, write term in full the first time it is used and give abbreviation in brackets); please note that the *Journal* does not use S for subject.

An abstract must be included. Although abstracts should be as tersely written as possible, they should summarize all the main experimental findings. The length of abstracts will vary with the length and complexity of the experiments reported but should usually be between 50 and 400 words.

An initial decision on a paper will normally be reached within 2 months of receipt. Where only minor revisions are requested, date of publication will be determined by date of receipt of the paper provided a satisfactory revision is received within 1 month of the request for a revision being made. The delay between final acceptance of a paper and its publication will normally be between 3 and 6 months. The date on which the manuscript is received in its final version (apart from minor alterations—see above) is printed.

Contributors receive 25 copies of their paper free. If they require further copies they should apply to the Publishers at the time of returning proofs.

The Quarterly Journal of Experimental Psychology is published for The Experimental Psychology Society by Academic Press, London and New York. Volume 23, 1971 (4 issues): subscription rate £6.50. plus £0.40. postage inland; $16.00 plus $1.00 postage abroad. Subscription orders should be addressed to Academic Press Inc. (London) Limited, Berkeley Square House, Berkeley Square, London W1X 6BA, with the exception of those originating in the U.S.A., Canada, Central America and South America; these should be sent to Academic Press Inc., 111, Fifth Avenue, New York, New York 10003. Application to mail at second-class postage rates at Jamaica, N.Y. 11431.

Undergraduates and graduates of less than three years' standing may obtain the journal at the reduced price of £3.60. They should apply initially to the Hon. Secretary, Dr David Legge, Department of Psychology, City of London Polytechnic, Central House, Whitechapel High Street, London, E1 79F. They should not send any remittance, but enclose suitable evidence of their student status or date of graduation, such as a letter from a university teacher or copy of a graduation list.

EDITORIAL

"Psychology's publication network appears among the most organized, vast, and envied systems of science" (Wolff, 1970): how pleasant to think that the *Quarterly Journal* is part of this magnificent system. The *Quarterly* is the official organ of the Experimental Psychology Society and must cater for the range of interest of its 217 members, each of whom gets his name into print at least once a year (see the end of this issue). The American Psychological Association publishes about a dozen journals which have become more or less specialized: despite the vast number of journal pages published by the APA (11,000 per year), the Experimental Psychology Society publishes 2·5 pages of journal per member per year whereas the APA only produces 0·3 of a page.

Publishing a journal that covers the whole range of experimental psychology makes it difficult to have any clear editorial policy. It also imposes a special duty on contributors to write in such a way that they can be understood by readers who are not specialists in a particular field. The *Quarterly* serves two purposes not served by most other journals. By juxtaposing articles on widely different topics it encourages the reader to browse outside his own speciality, and traditionally the Journal has allowed the writer more freedom to theorize in the context of an experimental article than is common in most of the more specialized journals. The present editors intend to continue this tradition. Although most of the articles published will continue to be reports on experiments, we hope also to publish theoretical and review articles on topics of current importance and interest.

Since the time when one man could adequately judge articles over the whole range of experimental psychology is past, we have adopted a system of associate editors. Potential contributions will be referred to the appropriate associate editor and the decision on their suitability will rest entirely with that editor. Our criteria for acceptance of articles are much the same as those in use elsewhere (see Wolff, 1970), and the most important consideration must be the article's contribution to the advancement of knowledge. We hope that intending contributors (particularly the young and experienced) will take the precaution of having articles vetted by colleagues for clarity of presentation before submitting: in our experience approximately 10% of articles submitted are so unintelligible that it is not possible to decide what experiment (if any) has been carried out. The most difficult papers on which to take editorial decisions are those that, while clearly written and reporting competently conducted experiments with enough subjects and good statistics, are nonetheless boring because the results are not surprising in themselves and throw no new light on theoretical issues. We feel the place for such contributions is in the more specialized journals and we shall try to avoid publishing them in the *Quarterly*.

Like most new editors we have made the usual good resolutions to speed up ecision time on submitted papers (normally within 2 months) and to get articles ublished fast (*between 3 and 6 months from receipt of the manuscript in its final* orm). We shall try to find referees *who are brisk but not brusque. Our book* reviews will combine candour with tact.

N. S. S.

Reference

WOLFF, W. M. (1970). *A study of criteria for journal* manuscripts. *American Psychologist*, **25**, 636–39.

Quarterly Journal of Experimental Psychology (1971) 23, 1–7

EVIDENCE OF ACOUSTIC CODING IN LONG-TERM MEMORY

H. C. A. DALE AND ALEX McGLAUGHLIN†
University of Hull

Contrary to previous indications, retroactive interference in long-term paired associate learning was found to be a function of acoustic similarity. Experimental groups were exposed to the A–B, A′–C paradigm where corresponding stimuli were homophones. Their retention scores were substantially and significantly lower than control groups run with an A–B, C–D paradigm. The failure of previous studies to reveal effects of acoustic similarity in this way is attributed to the use of an insufficiently high degree of similarity.

In recent studies of long-term memory (LTM) for verbal material, evidence regarding the coding procedures subjects employ has been obtained by varying item similarity in serial learning (SL), e.g. Baddeley (1966), and also by manipulating the similarity between the OL (original learning) and IL (interpolated learning) stimuli in paired associate learning (PAL), the dependent variable being OL retention. In SL, if an increase in item similarity along some specified dimension impairs performance it is inferred that the dimension which has been manipulated is one along which the items have been coded. A similar inference is made if interlist stimulus similarity along a specific dimension increases RI in PAL, e.g. Baddeley and Dale (1966), Dale and Baddeley (1969).

Both techniques have indicated that semantic coding is employed in LTM (Baddeley, 1966; Baddeley and Dale, 1966) whereas comparable manipulations of acoustic similarity have always failed to produce effects of any magnitude (Baddeley, 1966; Bruce and Murdock, 1968; Dale and Baddeley, 1969). In consequence, it has tentatively been concluded that LTM entails semantic coding but not acoustic. Bastian (1961) has pointed out that the kind of material employed in these investigations of semantic similarity is likely to be heavily contaminated with associative linkages. The distinction between semantic and associative factors is not at issue here so, for brevity, we will continue to refer to the operative dimension as semantic. The conclusion that LTM is coded semantically is at variance with the results of Brown and McNeil's (1966) retrieval study entitled "The tip of the tongue phenomenon". Their procedure, however, could have been biased since their questionnaire led subjects to generate acoustically related words. Evidence of acoustic coding has also been found from the application of a positive transfer paradigm in PAL by Sassenrath and Yonge (1967). They found transfer to List 2 pairs having stimuli which were homophones of those in List 1 was as great as

† The contribution of the junior author is to be offered in part fulfilment of the requirements of the Ph.D. degree in the University of Hull.

when the relevant stimuli were associatively related. This finding must be judged as no more than suggestive due to limitations of their design and the lack of clear statistical significance. The generalization, therefore, is not seriously contradicted. (Sassenrath and Yonge refer to homophones as homonyms, but we note that the *Concise Oxford Dictionary* is guilty of the same confusion.)

The technique of varying similarity along different dimensions and comparing its effects introduces a major methodological problem. In the absence of a universal objective measure of similarity, there is no way of ensuring that the degree of similarity along one dimension achieved with some particular set of experimental material is comparable with that achieved with a different set of material where a different dimension is varied. The failure to find effects of acoustic similarity in the studies quoted above could be attributed to a failure to achieve a sufficiently high degree of similarity in the experimental material, as was pointed out by Dale and Baddeley (1969). Some subjective justification for this supposition can be obtained from examination of the word lists used. Baddeley and Dale (1966) when varying semantic similarity, used synonymous adjectives given high similarity ratings in the lists prepared by Melton and Safier and published in Hilgard (1951). These pairs of words could be interchanged in general usage, viz. fearful–afraid, bragging–boastful. Comparable ratings of acoustically similar words are not available. The material used by Dale and Baddeley (1969) consisted of rhyming pairs, viz. fearful–cheerful, bragging–dragging. These pairs are not interchangeable when spoken. Thus the semantic and acoustic material that has been used can be seen to differ in its degree of similarity. The former achieved a level of equivalence which the latter did not.

The purpose of the present study was to seek evidence of acoustic coding in long-term PAL using the RI paradigm with homophones as the critical related stimuli With true homophones acoustic identity is achieved. Therefore the degree of similarity would be at least as great as for the semantic material of Baddeley and Dale (1966). In this way it was believed that a more valid comparison can be made with that study.

Experiment I

Method

All aspects of the design and procedure were closely modelled upon those used by Baddeley and Dale (1966) to ensure comparability with their study. Thus the subjects were required to learn two 8-pair lists in succession and were subsequently retested just once on the first one.

Design

In order to obviate any bias attributable to particular OL or IL lists, a counterbalanced design was employed in which each of two OL lists labelled A and X were coupled with each of two IL lists labelled B and Y. The four conditions derived from this design were as follows: 1—OL list A, IL list B; 2—OL list A, IL list Y; 3—OL list X, IL list B; 4—OL list X, IL list Y. The material is listed in Table I from which it can be seen that each stimulus in list A was matched with one in list B, the pair being homophones. Likewise the stimuli of lists X and Y formed matched pairs. Thus conditions 1 and 4 were experimental, OL and IL stimuli being similar, whereas conditions 2 and 3 were controls.

Material

The response words were all taken from the Baddeley and Dale (1966) study. The disyllabic homophones selected as stimuli were assigned to lists according to their frequency of usage in the language as assessed by the Thorndike–Lorge (1944) G-count. The aim was to balance the lists for word frequency, and this was roughly achieved. The stimuli were also vetted to minimize semantic and associative links as well as formal and visual similarity with any other stimulus or response. Difficulty was experienced in finding suitable material which would satisfy these constraints. It will be noticed from Table I that some of the homophonic stimulus pairs selected are slightly imperfect in that detectable acoustic differences would exist if they were pronounced very carefully.

Procedure

The group testing procedure of Baddeley and Dale (1966) was employed, but with visual presentation. For this purpose the material was printed in lower case letters on large white cards (3·1 in. by 5·5 in.) using rubber stamps. These cards were held up by the experimenter so as to be clearly visible to the whole group. The exposure was paced by a metronome.

Learning and test trials alternated, a freshly randomized order of items being used each time. For learning, each pair was presented for 4 sec. For testing, each stimulus alone was exposed for 4 sec. The responses were written in specially prepared booklets using one page for each trial. Each page had the digits 1 to 8 typed down the left-hand margin and to facilitate place-keeping the experimenter called out the ordinal number of each response during every test. At the end of each trial the subjects turned over the page so that their responses were hidden. The interval between presentation and test and that between trials was 10 sec. Both OL and IL lists were learned for 8 trials.

In the instructions given to the subjects the task was described as learning a set of passwords. When OL was completed they were told that the passwords had been discovered by the enemy and therefore needed to be replaced. At this point the OL response booklets were collected and new ones were supplied. After the eighth IL test the subjects were told to turn the booklets over and write the numbers 1 to 8 down the left-hand margin of the reverse side of the last page. As soon as they had all done this the retest of OL was announced and executed.

All 4 conditions were run twice, so a total of 8 groups was tested. This avoids the weakness of the intact group procedure with which some single accidental circumstance can contaminate the results of a complete group.

Subjects

Young enlisted men from the source used by Baddeley and Dale (1966).

TABLE I

Material used in Experiment I

List A		List B		List X		List Y	
Stimuli	Responses	Stimuli	Responses	Stimuli	Responses	Stimuli	Responses
mourning	creased	morning	lovely	rained	daring	reigned	common
bolder	feeble	boulder	recent	cited	swollen	sighted	narrow
byre	hollow	buyer	sinful	nightly	shabby	knightly	clever
conquer	tiny	conker	heated	corral	alike	choral	oily
wheeled	tuneful	wield	biting	studded	foamy	studied	windy
grater	funny	greater	fiery	eyry	coloured	eerie	shining
gamble	dizzy	gambol	certain	ceiling	flashy	sealing	equal
wasted	rusty	waisted	mixed	needing	stubborn	kneading	spicy

Results

For each subject, forgetting was assessed by the percentage loss, that is by subtracting the retest score from the OL score and expressing the difference as a percentage of the OL score. Raw scores were used as a measure and the index of OL performance was the score achieved on the eighth (final) trial.

The size of each group tested had not been completely under the control of the experimenter and prior matching was not possible. Consequently the total number of subjects tested in each condition varied from 28 to 36 and their OL scores varied from 1 to 8. As in the Baddeley and Dale (1966) study, OL performance was used as a guide for extracting matched groups of equal size for analysis. These groups ($N = 22$) were each constituted as follows: 12 subjects with a score of 8, 4 with 7, and 6 scoring from 4 to 6 inclusive. As before, subjects scoring 3 or less were discarded.

Details of OL, IL and retest performance are given in Table II. A 2-way analysis of variance was carried out on the percentage loss measures using OL list and similarity between OL and IL lists as the main factors. This revealed: (i) a highly significant effect of OL–IL similarity ($F = 11{\cdot}3$, $df = 1{,}84$, $P < 0{\cdot}01$); (ii) no significant effect of OL list ($F = 2{\cdot}8$, $df = 1{,}84$, $P > 0{\cdot}05$); and (iii) a significant effect of IL ($F = 5{\cdot}6$, $df = 1{,}84$, $P < 0{\cdot}025$) which, in this analysis, was the interaction term. Thus the retention of OL was significantly impaired by the interpolation of a list with acoustically similar stimuli.

TABLE II

Mean performance on the final OL and IL trials, together with retest scores and mean percentage loss

Condition	OL	IL	Retest	Per cent decrement
1	(A) 7·18	(B) 7·36	(A) 4·73	31·86
2	(A) 7·09	(Y) 7·45	(A) 6·64	6·31
3	(X) 7·14	(B) 7·45	(X) 5·55	24·43
4	(X) 7·18	(Y) 7·64	(X) 5·23	28·79
Mean experimental (Conditions 1 and 4)	7·18	7·50	4·98	30·33
Mean control (Conditions 2 and 3)	7·12	7·45	6·10	15·37

The particular list employed is denoted by the bracketed letter.

Experiment II

Although the overall results of Experiment I reveal a substantial effect of OL–IL similarity, their theoretical significance is weakened by the strong interaction term revealed in the analysis. To gain confidence that the main effects were reliable the experiment was re-run twice using new material and subjects from different

sources. These two runs will be referred to as Part A and Part B. The methodological details which follow are confined to the novel features of this experiment. The general method and the experimental design were the same as in Experiment I.

Material

A completely new set of monosyllabic words was introduced (see Table III). From the Table it can be seen that the lists were roughly balanced for stimulus frequency, as assessed by the Thorndike–Lorge count, although we note in retrospect that the American spelling of "tyre" is "tire". To prevent subjects coding by the initial letter of the stimuli, and thus defeating our intention that they should attend to the complete word, these were made redundant as can be seen from the table. All response words were of AA frequency.

Procedure

All material was presented using a Kodak Carousel slide projector. A timer controlled the duration of presentation. For Part A this was 4 sec, for Part B it was 3 sec. Since the changeover time was 1 sec the total cycle times were 5 and 4 sec, respectively. The interval between peresentation and test was filled with two slides carrying the messages "End of Presentation" and "Prepare for Test", respectively. That between the end of the test and the next presentation was also filled with two slides. The first read "End of Test"; the second, "Pencils down turn over". Only 4 learning and test trials were given with each list, which enabled a complete list to be run without a change of magazine.

Part A was run with intact groups having from 11 to 15 members. Part B was run twice using groups of 5 or 6 each time.

Subjects

Part A: students taking ONC courses at a College of Technology. Part B: first-year psychology undergraduates from Hull University.

TABLE III

Material used in Experiment II

List A		List B		List X		List Y	
Stimuli	Responses	Stimuli	Responses	Stimuli	Responses	Stimuli	Responses
tow (9)	came	toe (35)	down	wine (A)	came	whine (11)	down
tail (A)	part	tale (A)	quite	weak (A)	part	week (AA)	quite
tire (AA)	reach	tyre (4)	shape	weight (AA)	reach	wait (AA)	shape
taught (A)	kiss	taut (3)	land	where (AA)	kiss	wear (AA)	land
brake (23)	met	break (AA)	nice	heard (AA)	met	herd (37)	nice
bawl (6)	fish	ball (AA)	joy	hue (12)	fish	hew (10)	joy
beat (AA)	iron	beet (11)	gold	hymn (18)	iron	him (AA)	gold
boar (11)	view	bore (A)	yard	heal (27)	vote	heel (A)	yard

Frequency estimates of the stimuli taken from the Thorndike–Lorge 1944 G-count are given in brackets. All responses are rated AA.

Results

The data are summarized in Table IV. The coding in that table is the same as in Table II. Both parts of the experiment show substantial effects of OL–IL similarity. In Part A these were highly significant ($F = 15{\cdot}31$, $df = 1{,}43$,

$P < 0{\cdot}001$), whereas in Part B they were not ($F = 2{\cdot}4$, $df = 1{,}40$, $P > 0{\cdot}05$). In both experiments there were no significant effects of either OL or IL material. For these analyses, group size was equated with OL matching as in Experiment I.

TABLE IV

A summary of the results of Experiment II

	Condition	*N*	OL	IL	Rhtest	Per cent decrement†
	1	"	6·91	7·64	3·91	41·01
Part A	2	"	6·91	7·09	5·36	23·15
(technical college	3	"	6·82	7·36	5·82	16·39
students)	4	"	7·18	7·64	3·55	51·10
Mean exptl.			7·05	7·64	3·73	46·05
Mean contl.			6·87	7·23	5·59	19·77
	1	"	7·09	7·45	5·55	25·70
Part B	2	"	6·45	6·90	5·27	22·84
(university students)	3	"	7·64	7·45	5·82	25·22
	4	"	7·00	7·45	4·00	44·80
Mean exptl.			7·04	7·45	4·77	35·25
Mean contl.			7·04	7·17	5·54	24·03

† The figures in this column are calculated from each individual subject's decrement and not from the means shown in the OL and retest columns.

Discussion

Despite the failure to find a significant effect of similarity in Part B of Experiment II, the overall results of this study show that with homophonic stimulus pairs substantial effects of acoustic similarity can be demonstrated in LTM PAL using the RI paradigm. The combined results of both experiments show that the decrement with similar OL and IL stimuli is 190% of that shown by the control groups. This is in fact marginally greater than the 189% difference obtained in Baddeley and Dale's (1966) study of semantic similarity.

The discrepancy between performance on parts A and B of Experiment II and the differential effect of IL list on RI in Experiment I indicate the presence of a rather labile effect. The two groups used in Experiment II differed in academic status and it may be noteworthy that those of higher status were less influenced by acoustic similarity. It is possible that this signifies some relationship between academic ability and the strategy employed in our verbal learning task. Such a factor would not be unexpected for the results of semantic conditioning experiments show a similar relationship between ability and generalization along the acoustic dimension (cf. Luria and Vinogradova, 1959), but our supposition is no more than speculation.

Previous studies of acoustic coding in LTM have obtained results suggesting that some effect was present (e.g. Dale and Baddeley 1969; Dale and McGlaughlin.

1968). No doubt, had large enough samples of subjects been tested statistical significance at some conventional level could have been achieved. (We note, though, that Dale and Baddeley (1969) failed to produce significant results despite using an analysis which was biased in their favour.) In these studies, however, the effect was very weak compared with that obtained using semantic similarity (Baddeley and Dale, 1966). The reason for this is that a comparable level of similarity was not employed. The signal feature of homophonic pairs is that they provide a level of similarity which, at the point of identity, cannot be less than the level used in semantic similarity studies. The material used by Baddeley and Dale (1966) consisted of pairs of words which are interchangeable in speech. The acoustically similar words used in previous studies have fallen short of this criterion. Only with homophones is an equivalent level of similarity achieved and it would appear that this is necessary to obtain comparable effects of RI in PAL.

For providing the germinal idea underlying this study we are indebted to David Salter of Sheffield University.

The financial assistance of the Medical Research Council is gratefully acknowledged.

References

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology* **18**, 302–9

Baddeley, A. D. and Dale, H. C. A. (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. *Journal of Verbal Learning and Verbal Behaviour* **5**, 417–20.

Bastian, J. (1961). Associative factors in verbal transfer. *Journal of Experimental Psychology* **62**, 70–9.

Brown, R. and McNeil, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning and Verbal Behaviour* **5**, 325–37.

Bruce, D. and Murdock, B. B., Jr. (1968). Acoustic similarity effects in memory for paired associates. *Journal of Verbal Learning and Verbal Behaviour* **5**, 417–20.

Dale, H. C. A. and Baddeley, A. D. (1969). Acoustic similarity in long-term paired associate learning. *Psychonometric Science* **16**, 209–12.

Dale, H. C. A. and McGlaughlin, A. (1968). Axoustic similarity and retroactive interference in paired associate learning. *Psychonometric Science* **13** (4), 225–6.

Hilgard, E. R. (1951). Methods and procedures in the study of learning. In Stevens, S. S. (Ed.), *Handbook of Experimental Psychology*. New York: Wiley.

Luria, A. R. and Vinogradova, O. S. (1959). An objective investigation of the dynamics of semantic systems. *British Journal Psychology* **50**, 89–105.

Sassenrath, J. M. and Yonge, G. D. (1967). Meaning and transfer of verbal learning. *Journal of Educational Psychology*, **58**, 365–72.

Thorndike, E. L. and Lorge, I. (1944). *The Teacher's Wordbook of 30,000 Words.* Teachers College, Columbia University.

Received 10 *March* 1970

Quarterly Journal of Experimental Psychology (1971), **23**, 8—13

RECALL OF SEMANTIC CLUSTERS IN PRIMARY MEMORY

BETTY ANN LEVY† AND ALAN BADDELEY
University of Sussex

An experiment is reported which investigated the use of semantic relatedness as a retrieval cue in the primary memory component in a free recall task. Six-word semantically related clusters were placed in the middle and end positions of free recall lists. Retention was measured immediately after the list presentation and after a filled retention interval of 15 sec. Pure primary memory functions were calculated. The results indicated that semantic cues are useful in retrieval from secondary memory but played no part in recall from primary memory.

Introduction

Waugh and Norman (1965) made a distinction between memory for unrehearsed and recently input items or primary memory (PM), and memory for rehearsed items or secondary memory (SM). While PM and SM are independent systems an item may be resident in both stores. The capacity of PM is limited by the number of items it can hold. Tulving and Patterson (1968) demonstrated that while a set of related words was retrieved from SM as a single unit, the words were retrieved from PM as individual items. Thus the capacity of PM was not increased by the presence of related items in the terminal list positions. However, the probability of recall of the related words was substantially higher than the probability of recall of unrelated items occupying these same terminal list positions. It can therefore be asked whether the semantic or associative relatedness of the words acted as retrieval cues in PM, thus explaining the increased recall probability.

As PM is held to be an acoustic store, not receptive to semantic cues (Baddeley, 1966*a*,*b*; Levy and Murdock, 1968; Kintsch and Buschke, 1969) facilitation in the PM component in this free recall task, attributable to semantic cues, appears to be contrary to the notion that PM and SM employ different coding systems. Craik and Levy (1970), therefore, tested the hypothesis that what appears to be a facilitation in recall in PM is actually mediated by retrieval of these terminal items from SM. As the Waugh and Norman model allows for dual residence of the items (i.e. that an item may be in both PM and SM) a correction is necessary in order to obtain the pure PM function: thus the true probability of recall from primary memory (p_{PM}) is given by

$$p_{PM} = (p_R - p_{SM})/(1 - p_{SM}) \tag{1}$$

where p_{SM} is the probability of recall from secondary memory (estimated from performance in asymptotic list positions) and p_R is the total probability of recall at

† Visiting the University of Sussex on a Postdoctorate Fellowship from the National Research Council of Canada. Present address: McMaster University, Hamilton, Ontario, Canada.

each serial position (SP). Tulving and Patterson did not correct the recall in the recency positions for this dual residence factor. Craik and Levy (1970), however, estimated the probability of recall of related words in SM from lists containing the related set in the middle list positions. They then used this estimate to "correct" the PM recall of lists in which the related items occurred in terminal list positions. Using this procedure the pure PM component of lists with related items in the final serial positions was, in fact, smaller than the pure PM component of control lists. Thus the semantic cues did not facilitate free recall from PM, but appeared to be detrimental to recall.

The present study attempted to provide a more direct test of the suggestion that related items are resident in SM, though they occupy terminal list positions. According to Glanzer and Cunitz (1966) the PM component of memory can be eliminated by inserting a filled delay interval before recall of the list. The delayed recall curve does not contain the recency effect found in immediate recall. Delayed recall, thus, shows the probability of recall from SM after the list presentation. Baddeley (1970) has provided a formula for estimating the true contents of PM using this delayed recall procedure by the formula

$$N(I - D)/(N - D) \qquad (2)$$

where I is the probability of immediate recall, D is the probability of recall after a delay and N is the number of items in the list. The Baddeley estimate is mathematically like the Waugh and Norman (1965) estimate except that it is summed across serial position to give an estimate of PM capacity, rather than PM recall at each serial position.

The present experiment, therefore, measured the recall of related and unrelated words, occupying middle and terminal list positions, both immediately after the list presentation and after a filled interval of 15 sec. The raw scores of the delayed recall condition should provide direct evidence of the registration of the terminal list items in SM. Primary memory capacity can be estimated using the Baddeley (1970) formula.

Method

Subjects

The subjects were 24 student volunteers from the University of Sussex. They were paid for their participation in the experiment.

Material

Two sets of 18 lists were used in the experiment. There were 20 words per list. Each set of 18 lists was constructed from a pool of 288 unrelated words plus twelve 6-word semantic clusters. Each pool of 288 unrelated words was a randomly drawn subset of a larger pool of 456 unrelated words used by Craik and Levy (1970). The twelve 6-word semantic clusters consisted of 6 medium frequency members of a category (Battig and Montague, 1969), such as "onion, spinach, cabbage, raddish, turnip, lettuce". The lists in each set were of three types: control lists containing 20 unrelated words drawn randomly and without replacement from the pool of 288 unrelated words; "semantic middle" lists in which a 6-word cluster occupied serial positions 6 to 11 and "semantic end" lists in which the cluster occurred in serial positions 15 to 20. Unrelated words from the pool of 288 filled the other positions of the clustered lists. Those clusters of the pool of 12 which

acted as "middle" clusters for one set of material, were used as "end" clusters for the other set of 18 lists. The three types of lists were not blocked, but occurred randomly within the sets of 18 lists.

Procedure

The experiment was conducted using 4 groups of 6 subjects each. Thus two groups of subjects received each of the two sets of lists. Three lists of each type (semantic end, semantic middle, control) were recalled immediately after presentation and 3 lists after a filled interval of 15 sec. Those 9 lists which were recalled immediately for one group of subjects using each set of 18 lists, were recalled after a delay for the second group of subjects (and vice versa). Subjects were not told in advance which lists were to be recalled immediately and which lists after a delay.

The words were read by the experimenter at a rate of 1 word/sec. After the last word subjects received either the recall signal (to begin recall immediately) or the experimenter began to read a list of two digit numbers (one number/sec). The subject's task was to add 1 to each number read and to write the sum in the booklet provided. After 15 such numbers the recall signal was given. Recall of the lists was written, with a 1-min interval for recall of each list. The instructions were those of free recall with the exception that subjects were asked to recall first as many words as they could remember from the end of the list, then write down the earlier words, thus ensuring that the terminal words are retrieved from PM and the earlier words from SM, as is the case in standard free recall.

Results

The results of the present study can best be discussed with reference to Figure 1, giving serial position curves for the three types of lists at each of the two retention intervals. The serial position curves obtained in immediate recall replicate those found by Craik and Levy (1970). As they pointed out, the fact that the control condition yields the standard free recall curve, despite the slight modification in the instructions, suggests that the data are representative of free recall studies. Further, the recency effects found in all conditions verify that the instruction to recall from the end first was followed by the subjects. The curves obtained resemble each other in all positions except those containing the cluster words. The "middle cluster" of the semantic middle lists (SPs 6–11) yielded significantly higher recall than the control words in those positions ($t = 11{\cdot}69$, $df = 23$, $P < 0{\cdot}001$) and the "end cluster" recall (SPs 15–20) was superior to its control recall ($t = 5{\cdot}05$, $df = 23$, $P < 0{\cdot}001$). The results, then, replicates the original findings reported with standard free recall (Tulving & Patterson, 1968).

The results of the delayed recall condition are those of most interest to this paper as they provide direct evidence of registration in SM. The serial position curves obtained for the control and semantic middle lists lack the recency effect found with immediate recall, as would be expected if the PM component is eliminated by delayed recall (Glanzer and Cunitz, 1966). Again the middle cluster of the semantic middle lists gave significantly higher recall than the control lists, in SPs 6–11 ($t = 8{\cdot}59$, $df = 23$, $P < 0{\cdot}001$). The semantic-end lists, however, did not yield the standard delayed recall curve. This condition shows a strong recency effect even after the 15-sec delay. The semantic cluster in the recency positions gave recall scores equal to those found for the semantic cluster in the middle list positions (3·40 words per cluster for the semantic end lists and 3·07 words per cluster for the semantic middle lists). The recall of the semantic

cluster in SPs 15–20 was significantly superior to the recall of control words in these positions ($t = 6{\cdot}59$, $df = 23$, $P < 0{\cdot}001$).

The results of the semantic end lists in delayed recall give clear cut evidence that some cluster words are resident in SM, although they occupy recency list positions. It is therefore of interest to look at the pure PM functions for each type of list to see if the dual residence factor explains all the facilitation in the immediate recall curves. For each subject an estimate of SM recall was derived from recall performance at SPs 15 to 20 in delayed recall. Each subject's PM score, (SPs 15–20 in immediate recall) was then "corrected" by his SM score, according to the Baddeley (1970) formula. This gave estimates of PM capacity for each subject for each type of list. The median number of items recalled from PM was 3·00 for control lists, 3·18 for semantic middle lists and 2·21 for semantic end lists. Although there is a trend for fewer items to be recalled from PM for semantic end lists than for control lists, the effect missed statistical significance using a sign test ($Z = 1{\cdot}50$, $P > 0{\cdot}05$). Thus pure recall estimates of PM show no facilitation but perhaps a decrement in performance due to semantic relatedness.

Discussion

The present experiment attempted to provide evidence that the original recall facilitation in the recency position of the free recall curve, reported by Tulving and Patterson (1968), was mediated by secondary memory. That is, some items, though in terminal list positions, were in fact recalled from SM. The delayed recall results shown in Figure 1 offer strong support to this suggestion. Even after 15 sec delay the semantic cluster in the final serial position was recalled as well as the cluster in the middle serial positions suggesting that it is registered in SM. Further, the estimates of pure PM performance provided by the Baddeley correction procedure suggest that there is no facilitation in PM due to the semantic relatedness of the cluster words. Thus the data are not at variance with the suggestion that PM is an acoustic store in which semantic cues are not useful, as semantic relatedness did not facilitate recall in PM as it did in SM.

While agreeing in other details with the findings of Craik and Levy (1970), the recall decrement for semantic clusters in PM did not reach statistical significance in the present experiment. As the trend was in the direction of a semantic decrement, this failure to reach statistical significance may simply be due to the smaller number of subjects used in the present experiment. As Craik and Levy (1970) pointed out, this decrement could suggest semantic confusion effects occurring in PM but the explanation they favoured was that it resulted from a tendency for subjects to occasionally retrieve the cluster from SM, though it was also in PM. This would lead to a lower estimate of the contents of PM than was in fact the case. That the presence of the effect may be dependent on the number of subjects used in the experiment is not inconsistent with the Craik and Levy suggestion that the decrement may be an artifact of the subject's recall strategy.

Although Tulving and Patterson (1968) suggested that their original data were not reconcilable with two storage notions of free recall, preferring a two retrieval system explanation of the effect, it appears from the present paper that these data can be fitted into the PM/SM dichotomy of memory. Tulving and Patterson

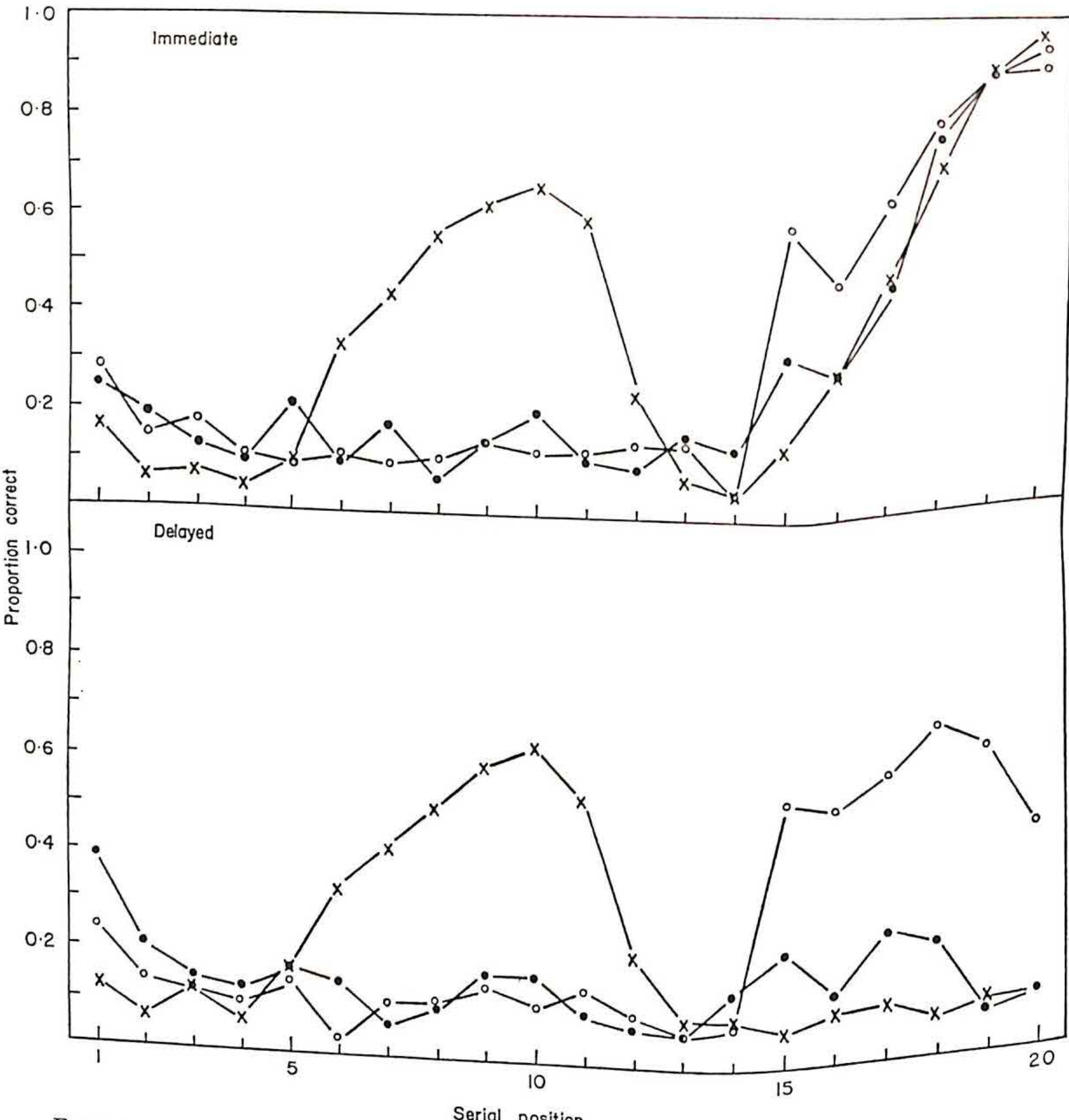

FIGURE 1. Serial position curves for three types of lists at two retention intervals. ●—●, Control; ×—×, semantic middle; ○—○, semantic end.

suggest that because related items are unitized in SM but not in PM "one must assume either that related words become unitized already in PM or STS, or that they are not unitized in PM or STS and become unitized in SM or LTS" (Tulving and Patterson, 1968, p. 246). The present results favour the latter alternative as related words appear to function as unrelated control items in PM. Favouring this explanation Tulving and Patterson maintain that one is faced with the necessity "to explain how related words not unitized in PM or STS are, almost without exception, transferred into, and how they then become unitized in SM or LTS" (Tulving and Patterson, 1968, p. 246). In the Tulving and Patterson experiment the probability of recall of the semantic cluster in the middle list

positions is, in fact, almost unity. In the present paper and in that of Craik and Levy (1970), however, the probability of recall of middle semantic clusters is about 0·50 to 0·60. This discrepancy in recall probability for the cluster in SM may be due to the differences in types of clusters used in these studies. Whereas the Tulving and Patterson clusters were either 4-word exhaustive categories or four high-frequency members of a non-exhaustive conceptual category the clusters in the later studies were 6 medium-frequency members of a category. Thus in the Tulving and Patterson situation it is possible to generate the remainder of the clusters given recall of part of that cluster, while the present clusters could not be generated but had to be remembered as members of an open class. It appears then that related words are transferred to SM without exception only if they comprise a set which can be generated completely given recall of one or more of its members. The problem of how items become unitized in long-term storage is not one unique to two-store theorists. Thus, the present data and that of previous "cluster" experiments appears to present little explanatory difficulty to two-storage theory, given that one allows that an item may be registered in both PM and SM at the same time.

This research was supported by a grant from the Medical Research Council to A. D. Baddeley. We are grateful to Dr F. I. M. Craik for helpful discussion.

References

BADDELEY, A. D. (1966*a*). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, **18**, 312–65.

BADDELEY, A. D. (1966*b*). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, **18**, 302–9.

BADDELEY, A. D. (1970). Estimating the short-term component in free recall. *British Journal of Psychology*, **61**, 13–5.

BATTIG, W. F. and MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: a replication and extension of the Connecticut category norms. *Journal Experimental Psychology Monographs*, **80**, No. 3, Part 2.

CRAIK, F. I. M. and LEVY, B. A. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology*, **86**, 77–82.

GLANZER, M. and CUNITZ, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behaviour*, **5**, 351–60.

KINTSCH, W. and BUSCHKE, H. (1969). Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, **80**, 403–7.

LEVY, B. A. and MURDOCH, B. B., Jr. (1968). The effects of delayed auditory feedback and intralist similarity in short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, **7**, 887–94.

TULVING, E. and PATTERSON, R. D. (1968). Functional units and retrieval processes in free recall. *Journal of Experimental Psychology*, **77**, 239–48.

WAUGH, N. C. and NORMAN, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89–104.

Received 5 *August* 1970

Quarterly Journal of Experimental Psychology (1971) 23, 14–21

ACOUSTIC AND ASSOCIATIVE CODING IN SHORT-TERM MEMORY

RICHARD E. LINDLEY AND DONALD R. BROWN
Psychology Department, California State College, Fullerton, Calif., U.S.A.

Short lists of word-digit pairs were presented to 456 college student subjects. One of the words was repeated as a memory probe either immediately after list presentation or after a short rehearsal interval. The stimulus words were either acoustically identical or associatively related (UP, DOWN). Both acoustic identity and associative relatedness produced a memory decrement which decreased with rehearsal. One interpretation of these results is that the primary memory trace is a multiple-dimension one and that, given time, subjects can recover non-acoustic information from it. The data also indicate that the "fate" over time is different for acoustically similar and associatively related items.

Introduction

An intriguing current notion about human memory is that there are two separate memory storage compartments (e.g. Adams, 1967; Waugh and Norman, 1965). One version of this notion is that, with either visual or aural presentation, verbal items are first coded acoustically and stored in a limited capacity (primary) memory system. With rehearsal, the coding shifts to a semantic base and the items are stored in a large capacity (secondary) memory system. When items are in the limited capacity system, it generally is acoustic similarity among the items which produces interference; when items are in the large capacity system, it is semantic similarity which produces interference (see Baddeley and Dale, 1966). While some semantic coding and interference may occur in primary memory, their effects are believed to be small when compared with acoustic effects.

The present study investigated the effects of acoustic similarity and associative relatedness using a short-term memory technique. The method of probed paired-associates (Murdock, 1961) was used. With this method, a short list of paired-associates is presented and then the response to one of the stimuli is probed by repeating a stimulus item as a recall cue. The response items were taken from a limited population whose members were known to the subjects (in this case, the digits 1 through 6). The critical stimulus items consisted of one, two, or three pairs of words in which the members of each word pair were either acoustically identical or associatively related. Take, for example, the stimulus-response pairs CITE-3, SITE-5, BEAR-2, BARE-6 with the probe CITE; (assuming the existence of proper controls) giving the response 5 instead of 3 would be an "acoustic" error and evidence for acoustic coding. Similarly, with the associatively related pairs UP-2, DOWN-1, GOOD-4, BAD-6 with the probe BAD, (assuming again the existence of proper controls) associative coding would be shown by the "associative" error of giving the response 4 instead of 6.

Using the short-term memory technique described above, the present study found evidence for both acoustic and associative coding. However, the distinction between an acoustic limited-capacity system and a semantic large-capacity system may still be valid since it is logically possible that the acoustic errors found in the present study represent interference from items still stored in the limited capacity system whereas the associative errors represent interference from items that have been already transferred from the limited capacity to the large capacity system. Such a hypothesis is capable of indirect test by interpolating an interval for rehearsal between the termination of the paired-associate list and the presentation of the probe stimulus; different predictions are made about the fate of the associative and acoustic errors during such an interval. The number of associative errors should increase with the interval as more items should be transferred by rehearsal to large capacity memory with a consequent increase in the possibilities for associative interference. On the other hand, with the homonyms used in the present study, there is no acoustic way of discriminating between them and hence the individual homonyms of a pair cannot be transferred to semantic large capacity memory by the accrual of meaning to the acoustic image of the homonym. Therefore, such rehearsal should, at best, leave intact those homonym-digit pairs available in limited capacity memory at the termination of list presentation or, more likely, lead to forgetting of these pairs because of interference from extraneous sources during rehearsal. This analysis leads to a different prediction about the effects of rehearsal on acoustically and associatively related items; viz. when subjects are permitted to rehearse associatively related word-digit pairs for a short period of time before the probe stimulus is presented, there should be an increase in the number of associative errors over the immediate probe condition because of the transfer of additional items to large capacity memory. On the other hand, rehearsal of homonym-digit pairs should produce either no change or more likey a decrease in the number of acoustic errors because of the inability to transfer these acoustically-coded items to semantic large capacity memory.

Method

Procedure

The same general procedure was used for all groups. The items were presented by a Kodak Carousel Projector with all time intervals controlled by the experimenter using a stopwatch. The time intervals described below include the approximately 0·7 sec for the carousel to change a slide. The stimulus items were words and the response items were the digits: 1 through 6. The words and digits were on separate slides and alternated as word, digit, word, digit, etc. The word "ready" signalled the start of each list and three asterisks signalled the end of the list; the asterisks were followed by one of the stimulus words which probed the appropriate serial position. The "Ready", word and digit slides were changed every 2·5 sec. For the groups with "Immediate Recall", the three-asterisk slide was presented for 2·5 sec and was followed by the probe stimulus slide for 10 sec during which time the subjects wrote down their digit answer and then covered it with a piece of paper. For the groups with "Delayed Recall", the three-asterisk slide was presented for 12·5 sec and then was followed by the probe stimulus slide for 10 sec. The subjects were instructed to rehearse the items during the 12·5-sec delay. The subjects were required to respond with a digit from 1 to 6 to every probe stimulus even if the response seemed to be a guess.

Experimental design

For eight of the groups tested, a $2 \times 2 \times 3$ factorial design was used in which one factor was time of testing (Immediate vs. Delayed Recall), another factor was type of lists (Homonym vs. Control Word Stimuli), and the third factor was length of lists (2, 4 or 6 paired associates). Time of Testing was a between subjects factor since subjects were tested with either the immediate or delayed recall procedure and Type of Lists and Length of Lists were within subjects factors since subjects were tested with both homonym-digit (H) and control word-digit lists (X) and with all three lengths of lists. For four of the groups, the H lists were presented first and were followed by the X lists; for the other four groups, the order was X lists followed by H lists.

Another eight groups of subjects were tested in an analogous fashion with associatively related word-digit (A) and their control word-digit (Y) lists. The $2 \times 2 \times 3$ factorial design had a between subjects factor of time of testing (Immediate vs. Delayed Recall) and within subjects factors of type of lists (Associative vs. Control Word Stimuli) and lengths of lists (2, 4, 6 paired associates). The A lists were presented first to four of the groups and followed by Y lists; for four other groups, the order was Y lists followed by A.

The three different lengths of lists were not used equally often; instead for the H, A, X, Y lists, there were two lists of length two paired-associates, four lists of length four, and six lists of length six. Each serial position within the lists of length 2, 4, 6 was probed once.

Materials

The words in the H lists were matched with the words in the X lists, and the A words were matched with the Y words on the basis of length, initial letter, and Thorndike and Lorge (1944) G word frequency. Experimenter's judgments were used to classify words as homonyms and associatively related. The associatively related word pairs were chosen to be opposites in meaning (antonym). Two different sets of H lists (H_1, H_2) were prepared as well as X_1, X_2, A_1, A_2, Y_1 and Y_2 lists. The sets of lists differed in the particular word-digit pairings and the stimulus words which comprised particular lists.

The way in which the lists were constructed is illustrated in Table I. For the H_1 lists (one of which is shown in Table I), the selection of the digits from 1 to 6 to be paired with a particular homonym, the order in which the lists of different length were presented, and the particular serial position to be probed within a given list was determined randomly without replacement. Such determinations then fixed the same parameters for the X lists (see Table I for an example). The same procedure was used for constructing the H_2, X_2,

TABLE I

Sample lists used

	From series H_1	From series X_1	From series A_1	From series Y_1
	HEAL-5 HEEL-6 BLEW-3 MAUL-1 BLUE-4 MALL-2	HUSH-5 HATE-6 BULL-3 MUSK-1 BOOK-4 MULL-2	FOUND-5 DOWN-1 SWEET-6 SOUR-2 LOST-3 UP-4	FLOOR-5 DROP-1 SCENE-6 SCAR-2 LAKE-3 US-4
PROBE	HEAL	HUSH	LOST	LAKE
CORRECT DIGIT	5	5	3	3
"ACOUSTIC" ERROR DIGIT / "ASSOCIATIVE" ERROR DIGIT	6	6	5	5

A_1, A_2, Y_1, Y_2 lists. The actual pairings of the lists learned by the subjects were: H_1X_2, X_2H_1, H_2X_1, X_1H_2; A_1Y_2, Y_2A_1, A_2Y_1, Y_1A_2. To illustrate, H_1X_2 means that a group of subjects learned lists H_1 followed by lists X_2. One group learned H_1X_2 with immediate recall while another group learned it with delayed recall; the same is true for all of the pairings shown.

Subjects

The subjects were 456 men and women college students from California State College, Fullerton, and they were tested in groups. There were 32 subjects in each of the H_1X_2, X_2H_1, H_2X_1 and X_1H_2 groups with immediate recall and 30 subjects in each of these groups with delayed recall. For the A_1Y_2, Y_2A_1, A_2Y_1, Y_1A_2 groups, there were 24 subjects in each with immediate recall and 28 in each with delayed recall.

Results

The homonym and control list data will be considered first. Each response to the homonyms and control words was classified as either a correct response, an acoustic error, or another error. Acoustic errors were illustrated in the Introduction of this paper and are defined as giving the digit in the list which was paired with the not-probed homonym. Of course, some of the responses scored as acoustic errors will be the result of guessing since the subjects were required to respond to every probe stimulus with a potentially correct digit. The control lists were included to assess the effects of such guessing; the scoring of acoustic errors for both the homonyms and their controls is illustrated in Table I. The percentages of correct responses, acoustic errors, and other errors are displayed in Table II. Inspection of this table reveals that the acoustic errors are in accord with the predictions made earlier, and analysis of variance shows that the differences are statistically significant. There are more acoustic errors with the homonyms than with the controls ($P < 0{\cdot}001$) because acoustic similarity produces interference. There was not a significant difference between immediate and delayed recall ($P < 0{\cdot}05$); however, the interaction: type of lists $\times$ time of recall was significant ($P < 0{\cdot}001$). The significant interaction comes about because, with delayed as compared with immediate recall, the number of acoustic errors with the homonym lists decreases while the number of such errors with the control lists increases. The decrease in acoustic errors for the homonyms with delayed recall is consistent with the notion that the homonyms were coded acoustically and were lost from limited capacity memory because they could not be coded semantically to enter large capacity memory. However, the correct response data are not in accord with this interpretation. Inspection of Table II shows the *decrease* in

TABLE II

Results for homonyms and control word lists (percentages)

	Immediate recall			Delayed recall		
	Correct response	Acoustic error	Other error	Correct response	Acoustic error	Other error
Condition:						
Homonyms	59·5	17·4	23·1	62·4	13·9	23·7
Control words	66·0	8·3	25·7	64·9	9·9	25·2

acoustic errors for the homonyms with delayed recall is accompanied by an *increase* in the percentage of correct responses. By the theory under test, the decrease in acoustic errors is due to the loss of items from limited capacity memory. Such a loss should not lead to improved recall but rather to a shift in errors from the acoustic to the other errors category. The fact that the rehearsal interval produced improved recall rather than a shift from one error category to the other suggests that the process occurring during such an interval is not interference and memory loss but rather the recovery of information which enhances memory. Since the homonyms are acoustically identical, the recovered information cannot be acoustic in nature. This, in turn, suggests that subjects were able to make effective use during the rehearsal interval of non-acoustic information in the memory traces of the homonyms.

The data of the associatively related lists and control lists will now be considered. The data for correct responses, associative errors, and other errors are presented in Table III. Associative errors were illustrated in the Introduction and are defined as giving the digit in the list which was paired with the not-probed associatively related word. The scoring of associative errors is illustrated in Table I.

Analysis of variance showed that the observed surplus of associative errors for the associatively related words over the controls was statistically significant ($P < 0{\cdot}001$). The interaction: type of lists × time of recall was not significant. Thus, associative relatedness adversely affected both immediate and delayed recall. Also, there were fewer associative errors with delayed than with immediate recall ($P < 0{\cdot}05$), for both the associatively related words and their controls. Inspection of Table III reveals that this observed decrease in associative errors with delayed recall is accompanied by an increase in correct responding. This finding supports the notion presented earlier that subjects can use the rehearsal interval to recover information to enhance memory. It should also be noted that an increase in associative errors with rehearsal predicted because of the transfer of additional items to large capacity memory was not obtained.

Another way of estimating the reliability of the present results is to compare separately the numbers of acoustic and associative errors in the lists of lengths 2, 4, 6 paired associates. For both the acoustically identical and associatively related items, there were more acoustic or associative errors in each length of list than for the respective control items. These relationships are shown in Figures 1 and 2 which also detail the serial position curves for the lists. Since the experimental design confounded items with serial positions, these curves should be interpreted cautiously.

TABLE III

Results for associatively related and control word lists (percentages)

	Immediate recall			Delayed recall		
	Correct response	Associative error	Other error	Correct response	Associative error	Other error
Antonyms	65·6	12·6	21·8	67·7	9·8	22·5
Control words	69·8	8·5	21·8	73·3	6·3	20·5

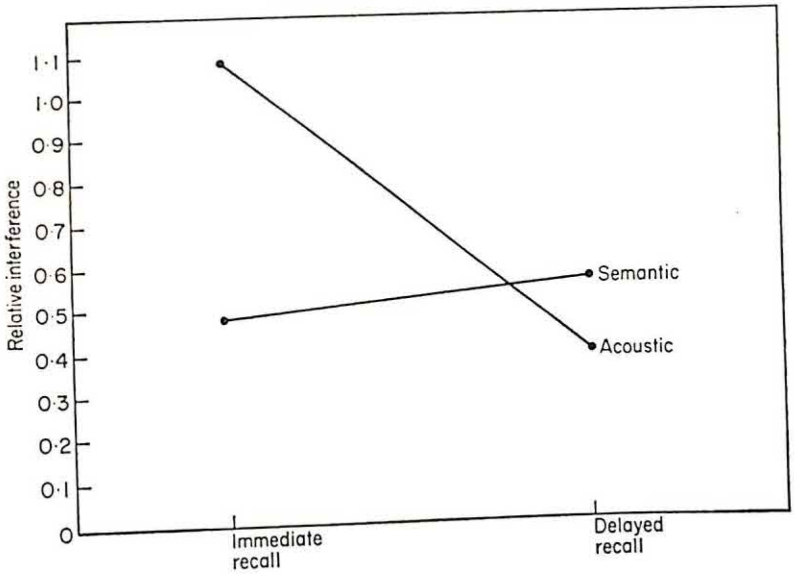

FIGURE 1. Proportion acoustic errors at each serial position for lists of length 2, 4, 6 paired associates.

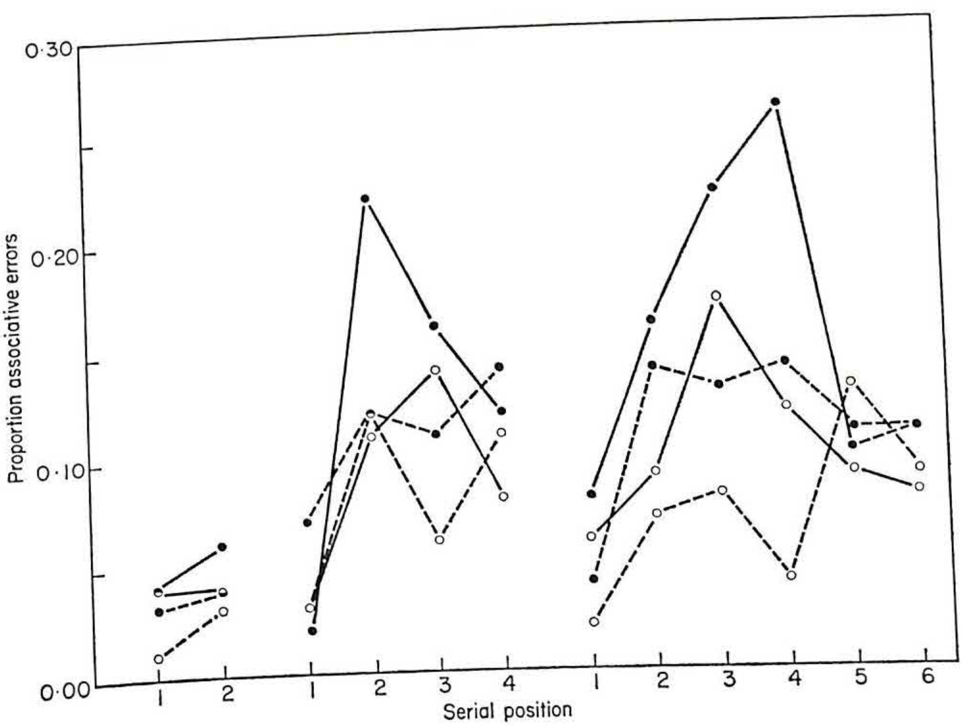

FIGURE 2. Proportion associative errors at each serial position for lists of length 2, 4, 6 paired associates.

"Fate" of acoustically-similar and semantically-similar items

It is of interest to determine if the fate of memory traces from acoustically identical and associatively related items is the same over time. In order to compare the acoustic and associatively related items directly, a procedure analogous to the determination of relative retroactive inhibition was used. The proportion of *relative acoustic interference* was computed by subtracting the number of control word errors from the number of homonym errors and dividing by the number of control word errors. The analogous formula was used to compute relative associative interference. Figure 3 displays the relative acoustic and associativ interference for immediate and delayed recall. Figure 3 shows a large decrease in relative acoustic interference with delayed recall and a slight increase in relative associative interference with delay. Thus the effects of acoustic and associative similarity appear to be different.

Discussion

The homonym and control word data of the present study indicated that the adverse effects of acoustic similarity were temporary. A study by Kintsch and Buschke (1969) which compared the effects of acoustic (homonymous) and semantic (synonymous) similarity in short-term memory also found that the adverse effects of acoustic similarity were short-lived. However, the present study found, in addition, that the reduction in the adverse effects of acoustic similarity was accompanied by an increase in correct recall. Indeed, the Kintsch and Buschke study may reveal the same effect since their homonyms were recalled

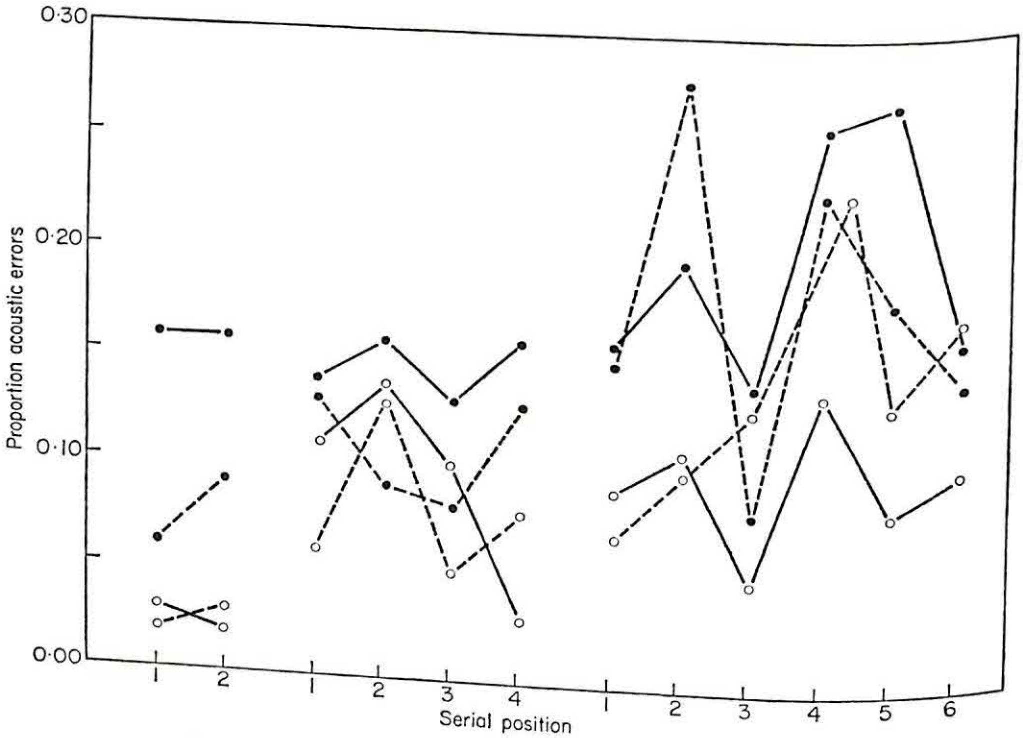

FIGURE 3. Relative acoustic and semantic interference.

at least as well as their control items once the initial adverse effects of acoustic similarity had occurred.

The temporary nature of the effects of acoustic similarity cannot be accounted for by a theory of primary and secondary memory in which coding in primary memory is exclusively acoustic. If only the acoustic properties of the homonyms are stored in primary memory, then recovery of non-acoustic information about the homonym pairs which might have aided discrimination and recall could not have occurred during the rehearsal interval of the present study. A modification of the theory permits it to account for the present data if it is assumed that non-acoustic associative factors play a more important role in primary memory than heretofore assumed. Alternately, it might be assumed that the items pass out of acoustic primary memory into semantic secondary memory during the 2·5/sec that each item was being presented. The alternate assumption requires, of course, that acoustic factors play a role in secondary memory.

The theory also assumed that rehearsal would lead to the transfer of more associatively related and control items to secondary memory. For the associatively related words, there should be a consequent increase in the possibilities for interference which would be reflected by an increase in the number of associative errors. However, the assumed increase was not obtained and, in fact, the number of associative errors for both the associatively related and control groups decreased with rehearsal. Such a finding can be readily explained if it is assumed that the associatively related and control words had already passed out of primary memory during the time these items were presented so that they were stored in secondary memory by the time the rehearsal interval began.

Finally, it should be clear that more data are needed to answer the questions raised by the present study. For example, the delay interval in the present study was unfilled and it has been assumed here that subjects followed instructions and used this interval to rehearse and recover information from the memory trace. This assumption should be checked by repeating the study with an interval filled by some rehearsal-decreasing activity.

This research was supported by a grant from the National Science Foundation to one of us (R.H.L.).

References

Adams, J. A. (1967). *Human Memory.* New York: McGraw-Hill.

Baddeley, A. D. and Dale, H. C. A. (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. *Jouranal of Verbal Learning and Verbal Behavior*, **5**, 417–20.

Kintsch, W. and Buschke, H. (1969). Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, **80**, 403–7.

Murdock, B. B., Jr. (1961). Short-term retention of single paired-associates. *Psychological Review*, **8**, 280.

Thorndike, E. L. and Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words.* New York: Columbia University, Bureau of Publications.

Waugh, N. C. and Norman, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89–104.

Received 13 April 1970

Quarterly Journal of Experimental Psychology (1971) 23, 22–33

MEMORY FOR SENTENCES AND NOUN PHRASES OF EXTREME DEPTH

CHARLES A. PERFETTI AND DOBA GOODMAN†

Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pa., U.S.A.

Memory for orally presented noun phrases and sentences was investigated in two experiments. Syntactic depth (D) of the phrases and sentences was varied with D approaching the values of 7 ± 2. Consistent with recent investigations, these studies found D not to be generally related to recall of sentences, phrases, or individual words within sentences. However, recall of individual words from an adverbial noun phrase apparently was related to D. The apparent relationship is accounted for not by depth of individual words, but in terms primarily of superior noun retention and its interaction with serial position effects. In addition, conditional recall probabilities for another type of phrase, a noun phrase with prenominal adjectives, indicate stronger relationships between adjectives and the noun than between successive adjectives. It is concluded that the Yngve grammar is deficient as a psycholinguistic model.

Introduction

In previous investigations of the depth hypothesis of Yngve (1960), the usual practice has been to employ sentences of moderate maximum depth (d_{max}), well under the values of 7 ± 2, which Yngve, taking the Miller (1956) magic number, suggested may represent the maximum depth allowed in the language. Studies reporting evidence in support of the hypothesis (Martin and Roberts, 1966; Martin, Roberts, and Collins, 1968) and studies reporting negative results (Martin and Roberts, 1967; Perfetti, 1969*a,b*) have both used $\bar{d}$, the average number of left branches in the phrase structure tree diagram, as depth measures. Since $\bar{d}$ and d_{max} are closely related, the question is not so much which measure is used, but rather whether one should expect memory to be affected by sentences of only moderate depth. High-depth sentences used by the present investigator have generally been less than $\bar{d} = 2{\cdot}00$ with a d_{max} of 3 or 4. However, in view of the findings that $\bar{d}$ is unrelated to retention, we were interested in whether even extreme depth affects sentence retention.‡

The rationale of the depth hypothesis has been discussed elsewhere (e.g. Yngve, 1960; Martin *et al.*, 1968; Perfetti, 1969*a,b*), and therefore can be omitted here. Two experiments are reported below. In the first, sentences of extreme depth and consequently of extreme length, were presented for tests of immediate memory. In the second experiment, adverbial noun phrases and adjectival noun phrases,

† Present address: York University, Toronto, Ontario, Canada.

‡ Some research (e.g. Tannenbaum, Wood and Williams, 1968) has been directed at the encoding aspects of the Yngve grammar, but we are concerned here strictly with decoding aspects.

the former being of the type responsible for the high-depth sentences of Experiment I, were tested for retention after 10 sec. For both experiments the question was this: what happens to memory for constructions which *approach* the limits of depth ($d_{max} = 7 \pm 2$) which Yngve hypothesized to place severe limitations on memory? The word *approach* is emphasized because it is virtually impossible, as Yngve observed, to construct sentences beyond $d_{max} = 7$ that are readily interpretable as English.

Experiment I

The first experiment concerned the question of how the retention of a sentence with $d_{max} = 7$ compared with the retention of the same sentence when its d_{max} was reduced to two or three. An example of a high depth (D) sentence is the following:

> *When a clearly less than carefully concealed weapon was noticed by detectives, the man was arrested.*

This sentence, which has $d_{max} = 7$ at the word *clearly*), can be reduced to $d_{max} = 3$ by transforming it to the following:

> *The man was arrested when detectives noticed a weapon which clearly was less than carefully concealed.*

A comparison of the recall of the sentence under these two conditions provides a test of the possible memory limitations imposed by extreme depth. In addition, the pattern of recall errors within the adverbial noun phrase (ANP) should also be revealing, since one interpretation of the Yngve model suggests that errors should be more frequent at the point of maximum depth, e.g. at the word *clearly* in the high D version of the sentence.

Method

Sentences

Twenty sentences were constructed with $d_{max} = 7$. This extreme depth was introduced by a subordinate clause containing an adverbial noun phrase (ANP) of five or six words. The length of the sentences ranged from 14 to 19 words with a mean length of 17·15 words. A low-depth version of each sentence was also constructed which retained the basic meaning and almost all the individual words of the high-depth version. Every word of the ANP appeared in the low-depth version. The length of these sentences ranged from 16 to 21 words with a mean length of 17·65 words. Their D-values were $d_{max} = 3$ for 18 sentences and $d_{max} = 2$ for two sentences. Both the high and low D versions were controlled for word frequency by reference to the Howes (1967) word count for spoken English. No more than one word per sentence was permitted that did not appear on the Howes list, and that word was retained on both the high and low D versions of the sentence.

In order that differences in D would not be confounded with one particular syntactic difference, various procedures were used to transform sentences from high D to low D. Some of these procedures were as follows: (a) a change in the order of the subordinate and main clause; (b) the insertion of relative pronouns to reduce the depth of the ANP; (c) a change in sentence form, e.g. from active to passive; (d) a change in word form, e.g. an adverb in the high D version was changed to an adjective in the low D version. Two lists were prepared from the resulting set of 20 high D and 20 low D sentences. Each list contained 10 high D and 10 low D sentences presented in alternating order.

Subjects and procedure

Forty University of Pittsburgh undergraduates participated in the experiment. There were 20 subjects for each list, 8 males and 12 females for List 1, and 10 males and 10 females for List 2. Test sentences and three warm-up sentences were presented via a tape recorder with written recall of the sentences required immediately after presentation. Subjects were tested individually.

Results and discussion

Since perfect recall of a sentence was seldom attained, the main analyses of the data are based on the probability of correct word recall. The recall scores treat as correct both transpositions in the order of words and the recall of a word with the same lexical root as the original word. However, the comparative results are unchanged when only exactly reproduced words are scored as correct.

Sentence data

The most straightforward test of the importance of D is a comparison of each high D sentence with its low D counterpart. The results of the comparison are clear: ten sentences showed higher recall in their high D version, and ten sentences showed higher recall in their low D version.

A simialr result was obtained when the data from the separate sentences and lists were combined. The average probability of correctly recalling a word was 0·72 for high D sentences and 0·72 for low D sentences. An analysis of variance confirmed that neither D nor list was a significant factor in recall, although the $D \times$ list interaction was significant ($F_{1,38} = 16{\cdot}3$, $P < 0{\cdot}01$).

One potential difficulty in interpreting these results is that the low D sentences were necessarily slightly longer than the high D sentences, the mean lengths being 17·15 for high D and 17·65 for low D. While length presumably would affect measures of complete sentence recall, it appeared to have no effect on individual word recall: for the sentences of 14 to 16 words ($N = 14$), the probability of correct word recall was 0·70; for the middle range of 17 to 18 words ($N = 18$), $P = 0{\cdot}73$; and for the longest sentences of 19 to 21 words ($N = 8$), $P = 0{\cdot}71$.

These data thus indicate that D does not affect the probability of recalling a word from the entire sentence. However, it is reasonable to suppose that the effect of depth might not extend over the whole sentence, but rather may localize itself in that part of the sentence which contains the high D construction.

ANP data

It is the adverbial noun phrase, functioning as the subject of a subordinate clause, which makes the crucial contribution to the depth of the high D sentences. It is therefore appropriate, having rejected the hypothesis that D affects the overall recall of words from the entire sentence, to examine what happened within these phrases.

The first fact worth noting is that the ANP was the most difficult part of the sentence to remember, regardless of depth. The probability of the correct recall of a word from the entire sentence was 0·72, whereas the probability of the correct recall of a word from the ANP was about 0·57. This difference is also apparent

in Figure 1, which shows errors as a function of sentence position. The abscissa indicates each of the first five words of the sentence and each of the last five words of the sentence, with M representing the middle sentence positions from 6 to L-5, regardless of sentence length. In general, for the high *D* sentences, the first five positions include words from the ANP, while for the low *D* sentences words from the ANP are included in the last five words. Since the *M* position includes words from the ANP for both types of sentence, the strong effect of the ANP is actually somewhat attenuated by the curves, but the main point is clear: the errors follow the ANP whether it is early in the sentence or late.

A comparison between the ANP phrases of the high *D* sentences with the corresponding phrases in low *D* sentences showed that in 12 out of the 20 compari-

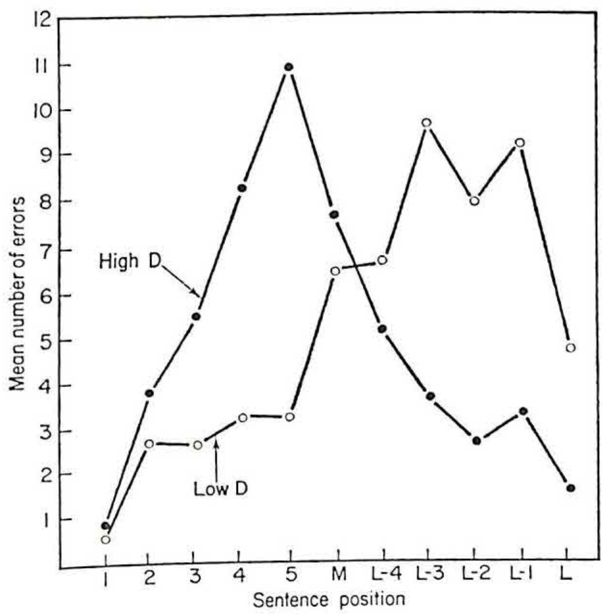

FIGURE 1. The relationship between recall errors and relative sentence position (Experiment I)

sons the recall of individual words was better for the low *D* phrase. The mean probabilities of correct word recall were, in fact, 0·565 for the high *D* phrase and 0·580 for the low *D* phrase, a non-significant difference ($F < 1$). Corroboration was provided by an analysis of the probability of a completely correct recall of the entire ANP. The mean probabilities, combining the data from the two lists, were 0·085 for the high *D* phrase and 0·070 for the low *D* phrase. An analysis of variance showed no significant main effects or interactions.

The picture that emerges from these measures fails to substantiate the hypothesis that a high depth adverbial noun phrase is more poorly retained than an adverbial noun phrase of lower depth. It is evident that extreme depth affects neither the recall of the sentence as a whole, nor the recall of the phrase in which it is located.

Position effects within the ANP

What happens to the retention of individual words as a function of their position within the ANP? Consider an example of an ANP from an actual sentence:

High D: *A clearly less than carefully concealed weapon. . .*

Low D: *A weapon (which) clearly (was) less than carefully concealed. . .*

When the recall of individual words from the ANP was compared under the two depth conditions, the following observations were made: (a) A word such as *clearly* was as well recalled when its D value was the maximum of seven as when its D value was reduced. (b) Words such as *carefully concealed* which occurred in the middle of the high D version were better recalled in the low D version where they occurred near the end of the phrase. (c) A noun such as *weapon* was well recalled in both its positions, although it was better recalled at the end of the high D phrase.

These observations suggest that the depth of a word is not an important factor, but that within the phrase there are significant primacy and recency effects, with the middle of the phrase being most vulnerable to error. However, such conclusions must be tentative because the ANP phrase was part of a long sentence. The findings are subjected to a more rigorous test in the next experiment.

Experiment II

Since extreme depth appears not to affect the recall of individual words from long sentences, the purpose of Experiment II was to determine whether it would affect the retention of short phrases occurring without any sentential context. The adverbial noun phrase, although rare, is the most typical high depth construction in English, and thus offers a straightforward test of this hypothesis. In addition, since Experiment I had shown the ANP construction to be the most difficult part of the sentence, regardless of depth, it was of interest to examine the source of errors within the ANP.

The experiment was also designed to test a further prediction of the Yngve model. The model analyses a noun phrase with prenominal adjectives as right branching, whereas the adverbial noun phrase is analysed as left branching. Thus, the syntactic difference between *a happy small child* and *a very small child* is that the former is analysed as *happy + small child*, while the latter is analysed as *very small + child*. Since right-branching structures are of a lesser depth than left-branching structures, the Yngve model predicts that the adjectival construction should be better retained in memory than the adverbial construction.

There were therefore three types of phrases used in this experiment: (1) a high depth adverbial noun phrase ($d_{max} = 5$); (2) the same phrase with reduced depth ($d_{max} = 2$ or 3); (3) an adjectival noun phrase with the same noun as (1) and (2) and having $d_{max} = 2$. These three phrase types will be referred to respectively as D_{max} (DM), D Reduced (DR), and Adjective Phrase (AP). Figure 2 shows an example of each type of phrase, together with the syntactic description assigned to them by the Yngve model. Thus, a direct comparison is possible between DM and DR, since they contain exactly the same words. The AP, of course, contains different words, but its depth is comparable to the DR phrase.

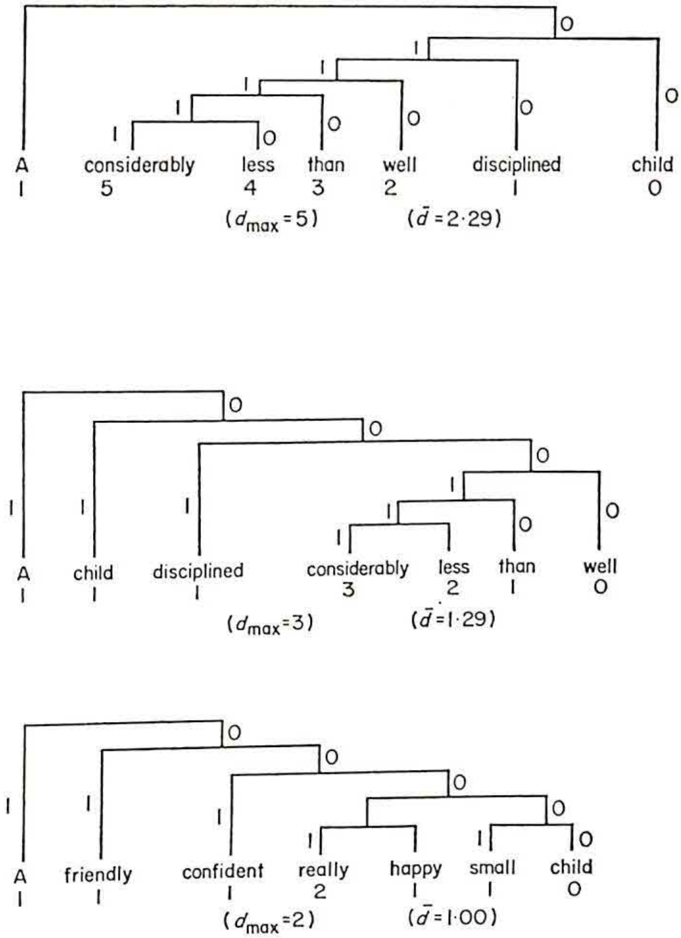

FIGURE 2. Syntactic analyses of the three phrase types according to the Yngve grammar.

Method

Phrase construction

In general, the DM adverbial noun phrases were taken from the sentences of Experiment I, although many were modified. DM phrases were rearranged, without the addition of any extra words, to construct the corresponding DR phrases. The AP was constructed by placing acceptable adjectives prior to the same nouns in the adverbial phrases.

One decision involving DR phrases should be noted. In normal usage, a DR phrase would typically include a relative pronoun and a copula verb. Thus, a normal version of the DR phrase in Fig. 1 might be *a child* who is *disciplined considerably less than well*. This lowers the depth of the phrase, but it also increases its length and decreases its lexical density. The potential confounding effect of these two variables suggested that the DR phrase should be constructed solely from words used in the DM phrase.

Design and procedure

Twenty-one nouns were employed in each of the three phrase types. Three lists of 21 phrases were constructed, each list containing seven representatives of each of the three phrase types in the same random order. Thus, the design varied phrase type within groups of subjects, but also allowed for a direct comparison of the recall of a given phrase between subjects; for example, the recall of *a considerably less than well disciplined child* which

appeared on List 1, could be compared with *a child disciplined considerably less than well*, which occurred in the corresponding position on List 2.

Phrases were presented on a tape recorder, and written recall followed a 10 sec interval during which the subject added a recorded series of seven numbers. A total of 60 undergraduates participated individually in the experiment, 20 for each list.

Results and discussions

Two measures were taken: (1) the probability of recalling an entire phrase, and (2) the probability of recalling an individual word from the phrase. (For both measures, only the exact reproduction of words was counted as correct.) The overall probabilities, based on the combined lists, are given in Table I for each of the three types of phrase.

The recall of DM and DR phrases

It is evident from Table I that there is unlikely to be any reliable difference in the completely correct recall of DM and DR phrases. The separate lists allow us to make 21 comparisons between these two types of phrase; of these, nine showed a superior recall of DR phrases, six showed a superior recall of DM phrases, and six showed no difference ($t < 1$, matched pairs test). It is a safe conclusion that depth did not affect total recall.

TABLE I

Recall probabilities for the three phrase types

Phrase type	Prob. completely correct	Prob. word correct
D Maximum	0·12	0·66
D Reduced	0·13	0·70
Adj. Phrase	0·10	0·70

The situation is less clear when the recall of individual words is examined. Of the 21 comparisons, 13 showed a superior recall of the words of the DR phrases, seven showed a superior recall of the words of the DM phrases, and one showed no difference. A *t*-test of the differences between the matched pairs was just significant on a one-tailed test ($t = 1{\cdot}71$, $P < 0{\cdot}05$). But an analysis of variance showed that within subjects the type of phrase did have a reliable effect upon retention ($F_{2,114} = 5{\cdot}52$, $P < 0{\cdot}01$); and the same result is obtained when the AP data are omitted ($F_{1,114} = 7{\cdot}74$, $P < 0{\cdot}01$). The between and within subjects analyses are brought into closer agreement, if the results of one particular phrase are omitted from the matched comparisons. Its DM version had 29% more words correctly recalled than its DR version; and since this accounted for almost 50% of the total sum obtained from those cases where DM recall was superior to DR recall, it is reasonable to regard this phrase as highly unrepresentative. Thus, it seems likely that extreme depth *does* affect the recall of individual words within a phrase.

Position effects

What happens to the recall of an individual word in a DM phrase, and how does the reduction of depth in the DR phrase affect recall? Figure 3, which shows the probability of a correct response for all three phrase types as a function of serial position, gives a partial answer.

The errors made in a DM phrase exceed the errors made for the DR phrase up until the last two words, where a reversal is observed. However, the difference is clearly greatest at position two, which was occupied by the noun in a DR phrase and by the first adverb in the DM phrase. This adverb typically appeared in position four of the DR phrase, and there, as can be seen from Figure 3, its retention probability was the same as in the DM phrase. This suggests that there is

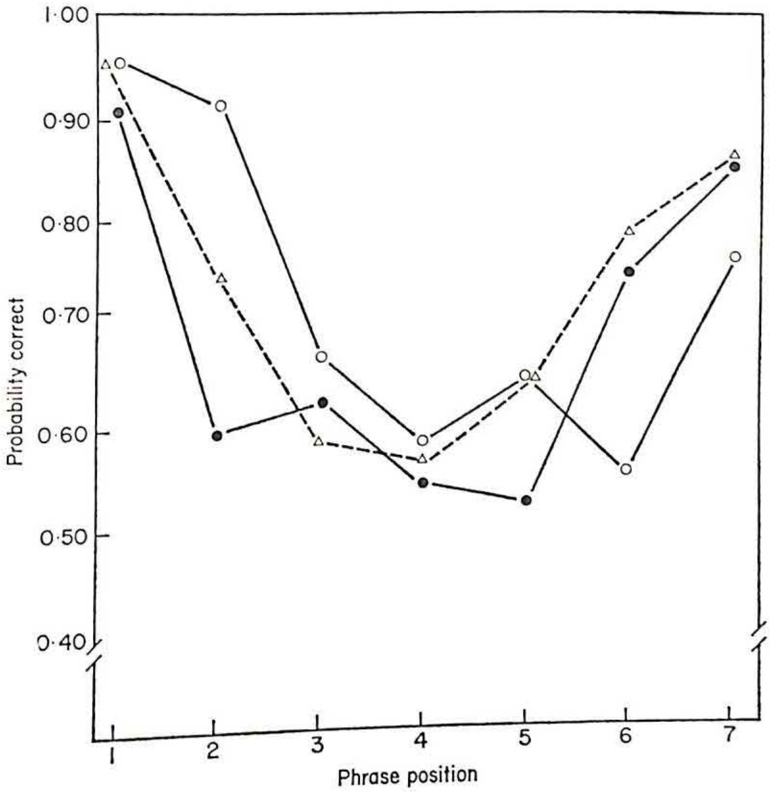

FIGURE 3. The probability of correct word recall as a function of position within a phrase (Experiment II). ●—● max D; ○—○, reduced D; △ - - - △, adjective phrase.

a primacy factor, a recency factor, and a noun factor. The primacy effect declines rapidly, but the noun factor is very powerful. Regardless of the noun's location, it is well recalled. When it occurs last, as in DM and AP, it is virtually at a ceiling and cannot be aided by recency. In DR, on the other hand, the final word is boosted in recall because it is not a noun and hence not well recalled in other positions.

The plausibility of this account is supported by an examination of the recall of individual words from the two phrases, as presented in Table II. The data are

for all phrases fitting the pattern shown by the words in the table. The words themselves are for example only.

By far the largest difference occurs for the fifth word of a DM phrase, which becomes the final word of a DR phrase. Although this difference corresponds to a difference in depth, it seems clear that the depth of individual words cannot serve as a general explanation of recall differences. This can be seen for the second, third and fourth words, where differences in depth do not relate to recall. The conclusions to be drawn are that subjects recall the noun best, regardless of its position, and that the remainder of the phrase is very difficult, regardless of the depth of individual words, with words in the middle of the phrase more difficult than words at the end.

The adjectival phrase

Comparisons between the AP and the other phrase types involve more than differences in syntax, since the AP is constructed from different words, apart from the noun. When the measure was the completely correct recall of a phrase, performance on the AP was lowest of the three phrase types (Table II), whereas when the measure was the recall of individual words, the AP ranked with the DR as the easiest. The reason for this is that adjective order, while it is not free in English, is less constrained than adverbial order. Thus, the measure of individual word recall shows an improvement for AP because the order of recall is irrelevant.

The position function for the AP resembles the type of serial effect obtained for unrelated words, or perhaps for words only sequentially associated (cf. the serial curve obtained by Deese and Kaufman (1957), for second-order approximation to English text). The last two points of the curve are very close to the points for the DM phrase, where both phrase types contain an adjective or adjective-like participle followed by a noun (*happy child* and *disciplined child*). It is interesting that recall for position two is poorer than for DR, but better than DM. This suggests that the noun is the primary storage unit, and recall of other words depends on their relatedness to the noun. Thus, an adjective in position 2 (AP) would be superior to an adverb in position two (DM) by virtue of its more direct relation to the noun.

TABLE II

Recall probabilities for individual words occurring in DM and DR phrases

Word	Phrase type DM	Phrase type DR	Recall Diff. DR-DM	Depth Diff. DM-DR
1. a	0·91	0·94	0·03	0
2. considerably	0·59	0·58	−0·01	2
3. less	0·62	0·52	−0·10	2
4. than	0·54	0·55	0·01	2
5. well	0·52	0·76	0·24	2
6. disciplined	0·74	0·66	−0·08	0
7. child	0·85	0·91	0·06	−1

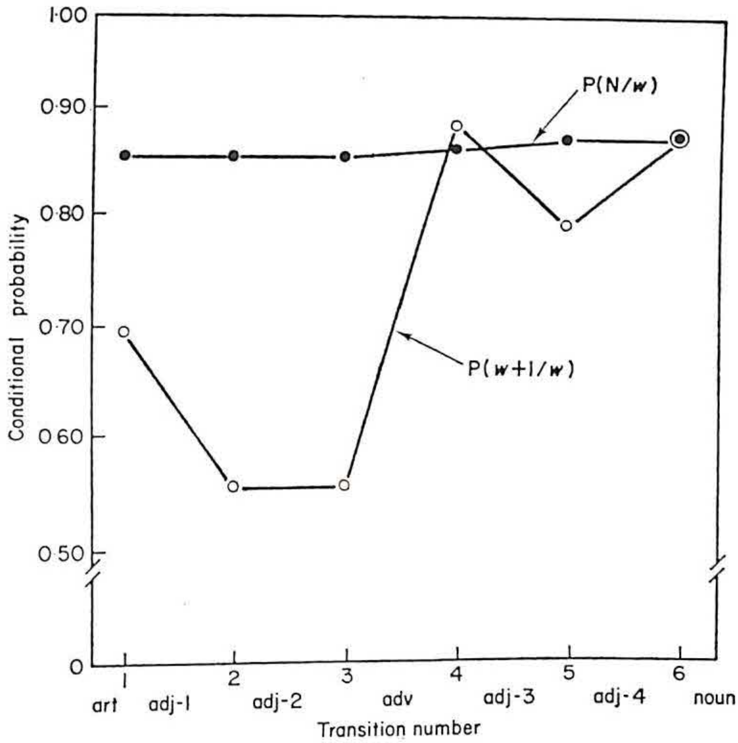

FIGURE 4. Conditional recall probabilities for the next word ($w + 1$) and for the noun (N) given correct recall of word w, for a noun phrase with prenominal adjectives (AP).

The way in which words of the AP are organized in memory may be elucidated by an examination of the conditional recall probabilities. Given that a word has been correctly recalled, what is the probability of correctly recalling (a) the next word and (b) the noun? Such probabilities may indicate whether adjectives are organized serially or each directly related to the noun. Figure 4 shows these conditional probabilities.

The main fact is that, given the correct recall of any word, w, the probability of correctly recalling the noun is higher than the probability of correctly recalling the next word, $w + 1$. This is the case for all the words except the adverb, which, as any grammar would suggest, is directly related to the adjective that follows. It is possible that the flat function for the conditional recall of nouns may be due to a ceiling effect, and that in a situation in which noun recall is poorer, one might not find the relationship shown here. However, it is a reasonable hypothesis that the recall of the noun is always high. In Experiment I, with very long sentences and recall which was *never* perfect, nouns were still correctly recalled over 80% of the time. This is also in agreement with the results of Mandler and Mandler (1964) who observed, in a serial anticipation task, very high noun recall more or less independent of the other words in the sentence.

For the data shown in Figure 4, the fact that the last three points of $P(w + 1/w)$ are so high indicates that the end of the phrase, for example, *really happy small child*, tended to be well recalled as a unit, with the weakest link occurring between the two adjectives. On the other hand, there appears to be little dependency

for all phrases fitting the pattern shown by the words in the table. The words themselves are for example only.

By far the largest difference occurs for the fifth word of a DM phrase, which becomes the final word of a DR phrase. Although this difference corresponds to a difference in depth, it seems clear that the depth of individual words cannot serve as a general explanation of recall differences. This can be seen for the second, third and fourth words, where differences in depth do not relate to recall. The conclusions to be drawn are that subjects recall the noun best, regardless of its position, and that the remainder of the phrase is very difficult, regardless of the depth of individual words, with words in the middle of the phrase more difficult than words at the end.

The adjectival phrase

Comparisons between the AP and the other phrase types involve more than differences in syntax, since the AP is constructed from different words, apart from the noun. When the measure was the completely correct recall of a phrase, performance on the AP was lowest of the three phrase types (Table II), whereas when the measure was the recall of individual words, the AP ranked with the DR as the easiest. The reason for this is that adjective order, while it is not free in English, is less constrained than adverbial order. Thus, the measure of individual word recall shows an improvement for AP because the order of recall is irrelevant.

The position function for the AP resembles the type of serial effect obtained for unrelated words, or perhaps for words only sequentially associated (cf. the serial curve obtained by Deese and Kaufman (1957), for second-order approximation to English text). The last two points of the curve are very close to the points for the DM phrase, where both phrase types contain an adjective or adjective-like participle followed by a noun (*happy child* and *disciplined child*). It is interesting that recall for position two is poorer than for DR, but better than DM. This suggests that the noun is the primary storage unit, and recall of other words depends on their relatedness to the noun. Thus, an adjective in position 2 (AP) would be superior to an adverb in position two (DM) by virtue of its more direct relation to the noun.

TABLE II

Recall probabilities for individual words occurring in DM and DR phrases

Word	Phrase type DM	Phrase type DR	Recall Diff. DR-DM	Depth Diff. DM-DR
1. a	0·91	0·94	0·03	0
2. considerably	0·59	0·58	−0·01	2
3. less	0·62	0·52	−0·10	2
4. than	0·54	0·55	0·01	2
5. well	0·52	0·76	0·24	2
6. disciplined	0·74	0·66	−0·08	0
7. child	0·85	0·91	0·06	−1

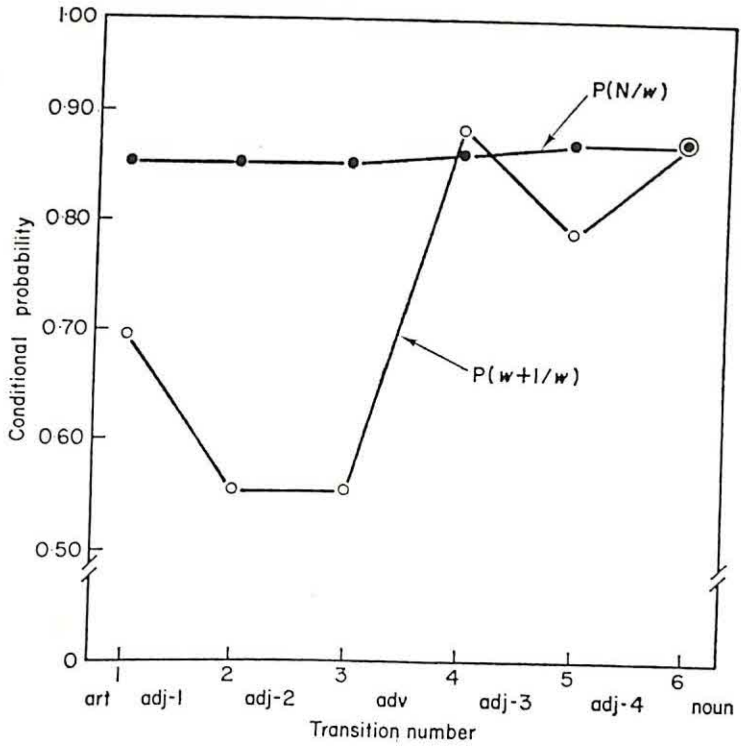

FIGURE 4. Conditional recall probabilities for the next word ($w + 1$) and for the noun (N) given correct recall of word w, for a noun phrase with prenominal adjectives (AP).

The way in which words of the AP are organized in memory may be elucidated by an examination of the conditional recall probabilities. Given that a word has been correctly recalled, what is the probability of correctly recalling (a) the next word and (b) the noun? Such probabilities may indicate whether adjectives are organized serially or each directly related to the noun. Figure 4 shows these conditional probabilities.

The main fact is that, given the correct recall of any word, w, the probability of correctly recalling the noun is higher than the probability of correctly recalling the next word, $w + 1$. This is the case for all the words except the adverb, which, as any grammar would suggest, is directly related to the adjective that follows. It is possible that the flat function for the conditional recall of nouns may be due to a ceiling effect, and that in a situation in which noun recall is poorer, one might not find the relationship shown here. However, it is a reasonable hypothesis that the recall of the noun is always high. In Experiment I, with very long sentences and recall which was *never* perfect, nouns were still correctly recalled over 80% of the time. This is also in agreement with the results of Mandler and Mandler (1964) who observed, in a serial anticipation task, very high noun recall more or less independent of the other words in the sentence.

For the data shown in Figure 4, the fact that the last three points of $P(w + 1/w)$ are so high indicates that the end of the phrase, for example, *really happy small child*, tended to be well recalled as a unit, with the weakest link occurring between the two adjectives. On the other hand, there appears to be little dependency

among, for example, *friendly*, *confidently*, *really* at the beginning of the phrase. The tentative conclusion is that adjectives tend to be directly associated with the noun rather than with each other. Syntactically, this seems to implicate a multiple branching phrase structure, i.e. branches from each adjective connecting to the noun, rather than exclusively binary branching. However, this can only be a tentative conclusion in the absence of any data on higher-order conditional probabilities.

The question of the psychological relatedness between noun and adjective appears to involve the syntactic ordering of the prenominal adjectives, which ordering in turn may be governed by lexical features. A study of cued recall by Lockhart and Martin (1968) showed that for noun phrases with two prenominal adjectives highly constrained in order, e. g. *the large red chair*, the second adjective was superior to the first as a retrieval cue for the noun. One interesting account (Martin, 1969) for this result is that lexical features such as definiteness or specificity of attribution are responsible for adjective ordering, with the more definite adjective being first in the underlying structure and, in English, transformed to the second position in the surface structure. However, it is not clear that such an account has a general application to adjective order, which, as in many of the phrases used here, often seems quite flexible.

One final datum may be of interest. To determine how subjects regarded the relative difficulty of the three types of noun phrase, we presented 21 trials, each composed of the three versions of the noun phrase, to 20 new subjects who ranked their order of recall difficulty. Their mean ranks from the three phrase types, from easiest to most difficult, were 1·29 for AP, 2·29 for DR, and 2·42 for DM. Approximately speaking, people appear to agree with the Yngve model in judging an adjective phrase as easiest, and an extreme *D* phrase as most difficult, although their rankings were not successful in predicting actual recall difficulty ($r = 0{\cdot}18$, $P > 0{\cdot}10$).

Conclusions

We have reported two studies, which taken together with previous research, suggest that depth is not a powerful factor in memory, in general. However, near the limits of depth suggested by Yngve, there is some effect, localized in the high *D* phrase. When such phrases are isolated from any sentential context, extreme depth does appear to reduce recall probability, not for the entire phrase, but for the first adverb of the phrase—the item in the position of greatest depth. However, the contrast between the high and low *D* phrases appears to be due not to the depth of individual words, but to the greater memorability of the last word in the phrase. The noun of the phrase appears to serve as the main item around which the phrase is organized, and the adjectives are more closely related to it than they are to each other, contrary to the serial organization implied by the Yngve model.

A final word about the Yngve grammar. It has continued to stimulate substantial interest because of its direct concern with the performance of the speaker and hearer and because of its exclusive concern with surface aspects of syntactic

structure. However, in addition to the linguistic deficiences of the Yngve grammar, there is now a considerable body of negative evidence to support the view that the study of psycholinguistic processes is not likely to advance by further use of the grammar as a general model. There is more to be gained by attention to the meaning of sentences, as a function of both their lexical items and underlying relationships.

References

HOWES, D. (1966). A word count of spoken English. *Journal of Verbal Learning and Verbal Behaviour*, **5**, 375–80.

LOCKHART, R. S. and MARTIN, J. E. (1969). Adjective order and the recall of adjective-noun triples. *Journal of Verbal Learning and Verbal Behavior*, **8**, 272–75.

MANDLER, G. and MANDLER, J. M. (1964). Serial position effects in sentences. *Journal of Verbal Learning and Verbal Behavior*, **3**, 195–202.

MARTIN, E. and ROBERTS, K. H. (1966). Grammatical factors in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, **5**, 211–18.

MARTIN, E. and ROBERTS, K. H. (1967). Sentence length and sentence retention in the free learning situation. *Psychonomic Science*, **8**, 535.

MARTIN, E., ROBERTS, K. H., and COLLINS, A. M. (1968). Short-term memory for sentences. *Journal of Verbal Learning and Verbal Behavior*, **7**, 560–6.

MARTIN, J. E. (1969). Some competence-process relationships in noun phrases with prenominal and postnominal adjectives. *Journal of Verbal Learning and Verbal Behavior*, **8**, 471–80.

PERFETTI, C. A. (1969*a*). Sentence retention and the depth hypothesis. *Journal of Verbal Learning and Verbal Behavior*, **8**, 101–4.

PERFETTI, C. A. (1969*b*). Lexical density and phrase structure depth as variables in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, **8**, 719–24.

ROBERTS, K. H. (1968). Grammatical and associative constraints in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, **7**, 1072–6.

TANNENBAUM, P. H., WOOD, B. S. and WILLIAMS, F. (1968). Effect of feedback availability upon generation of left- and right-branching sentence structure. *Language and Speech*, **11**, (Part 1), 12–19.

YNGVE, V. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, **140**, 444–66.

Received 15 *June* 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 34–45

ENCODING AND THE REA FOR SPEECH SIGNALS

MARK P. HAGGARD
Psychological Laboratory, Cambridge

A number of different possible explanations are distinguished for the findings on the right ear advantage (REA) for speech signals varied in their acoustic and phonetic properties. Two experiments are reported, using synthesized semivowels and vowels in monosyllable word frames. Both show REA. The detailed results of both experiments support the idea that a complicated "encoded" relationship between the acoustical stimulus and the response phoneme is a necessary condition for the REA, but that the encoding need only be signalled by an acoustical "trigger feature" in the stimulus; a task requiring a perceptual decoding is not necessary for REA to occur.

Introduction

Shankweiler and Studdert-Kennedy (1967, 1970) have demonstrated an advantage for the right ear (REA) in the dichotic identification of speech sounds differing only in respect of the segmental phonemes and the acoustical cues distinguishing them. This is a particularly profitable extension of the work of Kimura relating auditory perception to cerebral dominance (1961), in that detailed knowledge of the structure of the stimulus at the phonological, phonetic and acoustic levels can be employed in the attempt to explain the right ear advantage and the technology of synthetic speech can be employed. Exploration of the REA promises to say something about both interhemispheric relations and about the process of speech perception; but if it should be found that all speech sounds show REA then inferences to process may be limited. For example, bulky experiments would be needed to compare the magnitude of REAs in different conditions.

It has been suggested by Liberman, Cooper, Shankweiler and Studdert-Kennedy (1967) that encoding, a general property of speech sounds, is the essential factor in the experiments showing REA. "Encoding" refers to the principle whereby the phonemes are mapped onto the acoustical stream of speech in a fashion that is many-to-one, and one-to-many. Sometimes also called "non-invariance", this property arises from physiological and aerodynamic constraints in the production mechanisms, and from neurophysiological organisations that appear to have arisen to "combat" these constraints (MacNeilage, 1970). Liberman *et al.* further suggest that the perceptual "decoding" of an encoded signal requires a special mechanism located predominantly in the left hemisphere (for most normal right-handed subjects)—hence the REA. This inference from the experimental results involves some intermediate steps which have been considered by Darwin (1971) and other authors and will not delay us here.

Shankweiler and Studdert-Kennedy in two separate experiments (1967, 1970)

report an REA for stop-consonants but not for vowels. This is an interesting observation in the light of parallel findings on the phoneme boundary effect (Liberman *et al.*, 1967) but it is consistent with at least five plausible hypotheses about the necessary conditions for the REA:

(a) "*General conditions*" *hypothesis*

All speech sounds could potentially show REA but vowels lack some general prerequisite for its emergence not intrinsically connected with speech processing. Such a view has been suggested by Halwes (1969) on the grounds that dichotic vowels tend to fuse and not provide adequately separated and coherent inputs to the two hemispheres. If this were true the prospects of learning something about speech perception from dichotic experiments would be diminished.

(b) "*Acoustical processing*" *hypothesis*

REA depends primarily upon preferential processing in one hemisphere of certain *purely* acoustical properties, present in stop consonants but lacking in vowels.

(c) "*Trigger feature*" *hypothesis*

REA depends critically only upon the presence of an *encoded* acoustical feature in the stimulus, present in stop consonants and lacking in vowels. On this hypothesis the amount of "decoding" required in the perceptual task would not be important.

(d) "*Decoding*" *hypothesis*

REA depends upon the actual perceptual load of decoding presented to the subject, this being greater in stop consonants than in vowels.

(e) "*Response class*" *hypothesis*

REA depends only upon the phoneme class distinguishing the response items and not upon stimulus properties; vowels are among the classes not showing REA. While the distinctions between acoustical processing, trigger features, and decoding hypotheses are fairly subtle, these five hypotheses span substantially different conceptions of the REA. The present research was designed to assist the ruling out of four of them, it not being possible in this relatively new area of experimentation to regard any single experiment as crucial. An associated paper (Haggard and Parkinson, 1971) examines the role of a particular stimulus variable and of the task presented to the subject.

Encoding is a property of the stimulus but it is possible to distinguish it from other types of stimulus property. Encoding is not defined by any single acoustical correlate, but because the need for the encoding principle appears only with the need to move an articulator from one position to another the amount of encoding tends to be high when there is articulatory and hence acoustical change; it also increases with the speed of speaking. We can contrast this abstract property with simple acoustical properties such as presence of voiced vs. turbulent excitation, but in fact the distinction will become blurred in practice because articulatory changes are always accompanied by definable acoustical changes. These are the

changes in the frequency spectrum, known as "formant transitions". A vowel is characterized by a relatively stable (approximately 100 msec) set of values of *the resonant frequencies of the vocal tract, the formants. When a consonant is articulated these frequencies change rapidly (30 to 80 msec), assisting identifi-*cation of the consonant, but the values to which they go are relative to the neighbouring vowel, i.e. encoded. Hence the divorce of encoding as an explanatory principle from simple stimulus properties will depend upon use of a special form of encoding not involving formant transitions (Experiment II) and the demonstration that varying purely acoustical aspects of the formant transitions has no effect (Experiment I). Thus hypotheses (c) and (d) can be separated from (b); they can be separated from one another by determining whether REA depends upon decoding actually being required from the listener.

Experiment I

This experiment was designed to speak to hypotheses (b), (c), (d) and (e). It involved subjects identifying pairs of synchronized monosyllable words in interaural competition, where the distinction between the words was carried only by the initial consonant sound. The sounds used were the semivowels and laterals (hereafter semivowels) /w,r,l,j/. The last of these phonemes is the one usually rendered by "y" in orthography. Made with a relatively open vocal tract, these sounds have a short steady state which is acoustically like a vowel, but also have formant transitions leading into the neighbouring vowel. In the latter respect they are like stop-consonants, but the transitions are slower and more intense, being made with a more open vocal tract. Of all classes of speech sound the semivowels are those with which vowels have most in common; hence if they should show no REA, the phoneme class hypothesis would deserve some consideration; if no REA were obtained it would become unparsimonious although not altogether ruled out.

From the stimulus properties of formant transitions in themselves or as a hint of encoding in the stimulus we would predict an REA for semivowels; from the trigger feature hypothesis it would be of the same magnitude as that found for the place dimension of stop-consonants by Shankweiler and Studdert-Kennedy. Their other dimension, voicing, is not relevant to this class of sounds. However it is impractical to compare exact magnitudes, especially in experiments run under different circumstances, and an internal comparison is preferred for separating acoustical processing from trigger feature explanations. If REA depends solely upon an analyser of rapid stimulus changes present in stop-consonants and not in vowels, then normal semivowels should show only an intermediate REA, and semivowels with artificially speeded formant transitions should show larger REA. This can be directly tested.

Some light can be thrown upon the operation of actual decoding in the experiment by simply restricting the vowel context. Shankweiler and Studdert-Kennedy's subjects (1970) knew that the vowel could vary as well as the consonant, hence there was a real decoding problem. If a significant REA emerged with a *constant* vowel context, then the importance of actual decoding would appear diminished. The rationale of Experiment I is such that a result with no REA

weakly implies one alternative: the phoneme class hypothesis. Nothing can be inferred about the general conditions hypothesis, while the detailed pattern of a positive result would argue in favour of one of the others.

Method

Stimuli

The words "what, rot, lot, yacht" were prepared as digital data for the parallel formant synthesizer at Haskins Laboratories by manual retouching of the output of an algorithm which generates 95% intelligible speech (Haggard and Mattingly, 1968). A 40 msec steady state for the semivowel was followed by an 80 msec transition to the vowel. The vowel and final consonant were the same for all words, lasting 140 msec. The amplitude and pitch contours were identical, only the first three formant frequencies in the steady state and the consequent transitions distinguishing the utterances. A computer program was employed which laid down sequences of dichotic pairs on a stereo recorder, synchronized to better than 1 msec. A basic sequence of 120 trials was prepared which was subsequently spliced in the middle with the two halves interchanged to provide a further 120 trials. The intertrial interval was 5 sec and rest pauses occurred every 60 trials. The basic sequence of 120 items on one track was random without replacement; the sequence for the other track was random with respect to itself but constrained by the first such that the same word never occurred on the two tracks on one trial and the three pairings of each particular word occurred equally often. Two such 240 trial tapes were prepared, identical except that in the second the rate of transmission of the digital data to the speech synthesizer was speeded by a factor of two, halving all the above durations without any frequency multiplication. The inter-trial interval, however, remained 5 sec. In summary, the stimuli involved four words uniquely distinguished by initial semivowels, a fixed vowel context and the experimental treatment of speeding.

Subjects and procedure

Two groups were employed, receiving identical treatment except for the durations of the stimuli they heard. The subjects were normal-hearing right-handed undergraduates; large groups were employed to give some possibility of a difference between the groups emerging. The subjects heard the tapes via high quality stereo equipment and Sony DR3A earphones at 70 DB SPL (speech peaks) in testing groups of up to 18. Within each of the two experimental groups half the subjects began with a particular channel of the recording and reproduction equipment to a particular ear. Everyone reversed earphones at half-time. A 20-trial practice session was given to familiarize the subjects with the stimuli and the response task, which consisted of writing down the initial letters for the two words they thought they heard on each trial, placing the one about which they were more confident first. This procedure is the same as that employed by Shankweiler and Studdert-Kennedy. As might be expected, some subjects in the group hearing the speeded stimuli tended to hear "yacht" as "dot"; they were told that "yacht" was the correct response in this case, but in general the stimuli were acceptable, even when speeded, and the low scores are more the product of the dichotic situation than the stimuli.

Results

Only the results of the confident response are quoted, as in the group with the speeded stimuli performance on the other came fairly close to chance. Table I presents the confident response figures for the two groups. In the unspeeded group the effect amounts to about 5% REA, very similar to that found by Shankweiler and Studdert-Kennedy for the place dimension of stop-consonants. Comparison of the two groups suggests the acoustical factors of speed and duration of transitions are not relevant to the REA. The difference between the groups

TABLE I

Results of dichotic experiments on identification of /w, r, l, j/

Group = Stimulus condition	N	Mean percentage scores right ear	left ear	Right ear advantage	t	p
Normal	49	65.79	60.47	5.32%	3.429	<0.001
Speeded	37	61.95	58.71	3.24%	2.74	<0.01

Difference between advantages not significant, $t < 1\cdot00$.

is not itself significant, but is in the reverse direction to that required by the idea that REA is linked to the brief and rapidly changing acoustical characteristics of stop-consonants (acoustical processing).

Discussion

The result of this experiment suggests very strongly that properties shared between vowels and semivowels such as overall duration and openness of vocal tract are not relevant to the REA and that some properties shared by semivowels and stop-consonants such as the related presence of formant transitions and encoding are relevant. The sheer brevity of stop formant transitions appears not to underlie the difference found by Shankweiler and Studdert-Kennedy between vowels and consonants.

Thus the "phoneme-class" hypothesis is made somewhat unlikely by the finding of an REA for the class of phonemes most like vowels, but a demonstration of REA for vowels under some set of circumstances is required to rule it out. Finding an REA when there was no actual encoding in the situation favours the "trigger feature" hypothesis over "decoding" and finding that the most plausible acoustical factor is irrelevant favours trigger features over acoustical processing. Hence we may tentatively conclude that the perceptual mechanism is influenced by properties in the stimulus that signal a need for complex perceptual decoding. However this idea of a "trigger feature" is a novel one and a set of demonstrations parallel to those of Experiment I is needed to establish it. Also presence of some trigger feature is not itself a sufficient condition to generate an REA; the syllables used by Shankweiler and Studdert-Kennedy (1970) had consonant formant transitions but failed to show REA when the task was vowel identification. Hence the *task* is implicated, or rather a specific relation between the task and the critical stimulus variables; we would not expect formant transitions to act as the trigger feature and produce REA in the voicing distinction, for which they are not relevant, and the implications of this will be examined later. Meanwhile the need to place the "trigger feature" concept on a sound empirical basis plus some simple curiosity about Shankweiler and Studdert-Kennedy's failure to obtain REA for vowels produced Experiment II.

Experiment II

In general vowels are less context-dependent than consonants. However some perceptually significant encoding is observed in fast speech (Lindblom and

Studdert-Kennedy 1968); also vowel recognition depends upon the perceiver's assumptions about who the speaker is, about the size of his vocal tract and his characteristic articulatory positions (Ladefoged and Broadbent, 1957). As the degree of the former type of encoding operating in Shankweiler and Studdert-Kennedy's stimuli was uncertain, and the related variable of vowel duration has been shown to be irrelevant by Darwin (1969, 1971) it was decided to use speaker context rather than consonantal context as the source of encoding. It was hoped initially to employ a "mock" encoding as in Experiment I by having two speakers differing in pitch but not in mean values of the vowel formant frequencies. Although different pitches generate different assumptions in the perceiver about the speaker's vocal tract size the pitch itself does not come into vowel recognition as a direct cue. However, without attendant formant frequency differences pitch differences between synthetic "speakers" were found to produce differences in naturalness. Decreased naturalness could conceivably have underlain the lower REA for speeded semivowels in Experiment I, hence this arrangement would have been inconsistent with one of the purposes of the experiment—to obtain an REA if possible.

Accordingly a different approach was employed to distinguish between the importance of actual encoding in the experimental situation and potential encoding. Consider the case where a subject is instructed to attend to one ear only. Two major stimulus conditions can precede a given trial: the previous stimulus item could be in the same voice or a different voice. On trials preceded by a same-voice trial we can infer that the subject was tuned in not only to a particular ear but to the particular voice which he has just heard, if he is obeying the instructions and performing well above chance. Greater certainty that he was adjusted to the characteristics of this particular voice could be had by considering a particular trial only if the preceding trial (in the same voice) were correctly identified, but this would diminish the number of trials that could be scored. On these following-same-voice trials we could assume that the amount of decoding required is less if it is found that accuracy is greater than on other types of trial. Haggard (1966) has shown this is the case for frequency multiplication and division of speech— a transformation broadly similar to the major difference between speakers—vocal tract size. If an overall REA were obtained in a vowel identification experiment employing different speakers three possible explanations of its appearance there might be distinguished. If similarity of voice on the preceding trial had no effect, then the mere knowledge that there is more than one speaker must be operating, either via a long term expectation or a short term cue from changes between trials or differences between ears. If trials preceded by the same voice showed *greater* REA, then tuning *per se* to one type of voice would appear important. This would imply that voice difference augmented the effective separation between the two ear channels (Broadbent, 1962)—a possible physiological prerequisite of demonstrating ear differences. Then the lack of REA in Shankweiler and Studdert-Kennedy's experiment could be attributed to low effective ear separation for vowels. This "general conditions" hypothesis (a) could be further examined by inspection of intrusion errors; a greater relative proportion of intrusion errors on the left ear than the right would suggest that the REA interacted with attention

and ear-channel separation and did not reflect primarily some aspect of speech-processing.

Lastly, if trials preceded by a different voice showed greater REA then we could conclude that the actual encoding and hence decoding on a particular trial was important. Disregarding the complicated mechanism such a result would imply, it has a certain plausibility as an experimental prediction, and hence in the event of a null or opposite result we would wish to be sure that in the situation employed there was a genuine need for decoding, attested by an overall higher score for trials following same voice-trials.

Method

Stimuli

A stimulus tape was prepared in a fashion basically similar to that of Experiment I. The words employed were "a kid, a ked, a cad, a cud, a cod, a could", as specified in Table II. The sequence comprised 120 trials heard twice. Randomization was without replacements within the two 60-trial blocks of the basic sequence. In these blocks each word appeared twice with each of the five possible pairings with other words on the other ear. On one of these occasions the word was in a high voice and on the other in a low voice. On any one trial neither voices nor words were the same on the two ears. One-third of trials were preceded by the same voice, and among these low-low and high-high sequences were equally frequent. The words in the following same-voice trial were constrained to represent the six available ones as equally as possible on one track, the remaining selection being left to the randomization and the pairing principle mentioned.

Subjects and procedure

Forty normal-hearing right-handed undergraduates were randomly assigned to four groups to enable reversing of headphones within and across subjects as for Experiment I, plus order control for the instructions. These were radically different from those for Experiment I, and employed the attention paradigm of Kirstein and Shankweiler (1969). Here the subject is instructed to pre-attend and report from one ear only in a given block of the experiment. This attention is stressed and intrusions from the other ear are marked as errors. The authors of this technique hold that it gives basically similar results to the free recall technique, but its interpretation is obviously more satisfactory. In the present experiment, of the 20 subjects beginning the experiment with a given channel of the recording and reproduction equipment to a given ear, 10 attended to the left for the first block of 60 trials and 10 to the right. This is the variable "groups" in the analysis of the results. A symmetrical design was obtained by every subject changing ear of attention after trials 60 and 180, and reversing earphones after trial 120.

A 30-trial practice was given and no trouble was encountered with the non-existent response word "ked". The words were printed across the top of a specially laid out response sheet, and subjects had to place a tick in a column beneath the word they thought they heard on each trial. They were informed about the two ears never recieving same voice or item but were told nothing about the sequence of trials.

Results

This experiment produced average scores for the right and left ears of 58·7 and 53·4%, respectively, an REA of 5·6%. Analysis of variance showed this to be significant ($F = 20{\cdot}3$, $df = 1{,}36$, $P < 0{\cdot}001$). The sequential data were then obtained by marking separately the responses preceded by trials in the same voice. These scores were then multiplied by two to make them comparable with scores on trials by the same voice. Table III suggests, and the analysis of variance

TABLE II

Acoustical specifications of the stimuli for Experiment

Voice	Vowel formant frequency	Stimulus words					
		"kid"	"ked"	"cad"	"cud"	"cod"	"could"
High (onset fundamental frequency 164 Hz)	F1	286	489	640	640	537	386
	F2	1695	1695	1620	1155	769	769
	F3	2525	2525	2525	2525	2378	2378
Low (onset fundamental frequency 120 Hz)	F1	412	614	794	794	666	489
	F2	2156	2078	1845	1465	1455	1155
	F3	2694	2694	2694	2695	2525	2525

The durations of the four phones, in each case, e.g. /əkɪd/ (counting transitions within vowels) were 70, 80, 120 and 80 msec. There was a final release including transitions of 60 msec. In all cases the fundamental frequency (i.e. pitch) fell 10 Hz during the vowel, then dropped a further 20 Hz during the final /d/, remaining level at the low value through the release.

confirms (Table IV) that there is no dependence of REA upon similarity of voice on the preceding trial. This is found despite the significant advantage in having a sequence of two items in the same voice on a particular ear, and hence the presence of some need to "decode".

Discussion

The second experiment goes beyond the first in evaluating all of the initial hypotheses. The idea that REA is a property of a phoneme class and hence of the response task alone is ruled out by the finding that vowels can show an REA.

The sequential effects show that the amount of decoding actually required on a given trial is not relevant, and hence our trigger feature hypothesis is supported over the decoding hypothesis. A more detailed experiment by Darwin (1971) using different groups of subjects rather than sequential conditions has since produced the same conclusion among its results. In addition, an experiment on the perception of voicing distinctions reported by Haggard and Parkinson (1971) showed that the perceptual use of acoustical cues to the voicing of consonants is

TABLE III

Mean percentage scores in four sets of trials

		Ear	
		Right	Left
Voice on preceding trial	Same	61.9	55.6
	Different	57.1	51.8

TABLE IV

Summary of analysis of variance

Source	SOS	*df*	mean square	*F*	*P*
Total	24113	159			
Groups(G)	2349	3	783	1.90	NS
Preceding voice Same or different (V)	537	1	537	10.53	<0.01
Ears(E)	917	1	917	15.7	<0.001
G x V	181	3	60.3	1.19	NS
G x E	108	3	36	<1.00	
V x E	2	1	2	<1.00	
G x V x E	450	3	150	7.25	<0.01
Subjects (S) within groups	14881	36	413.5	20.00	<0.001
S x V	1844	36	50.9	2.46	<0.01
S x E	2097	36	58.3	2.82	<0.01
Residual (S x V x E)	747	36	20.7	—	

The scores for trials following trials in a different voice were raw, those following the same voice were doubled.

dependent upon the perceived place of articulation of a consonant—a form of simultaneous encoding. Oscillagraphic analysis of the particular acoustical cue used in that experiment has since shown that speakers do not *produce* contextual variations in it, although they do produce contextual variations in another cue that had to be present in the experimental stimuli with a fixed value. As in Experiment I, the sufficient conditions for emergence of REA in Haggard and Parkinson's situation appear to have been the requirement of a response involving a linguistic distinction, plus the presence of some normally encoded acoustical attribute relevant to that distinction. The encoding of the stimulus information actually usable for the distinction appears irrelevant. Hence several different experiments favour the "trigger feature" hypothesis over actual decoding.

It will be noticed that Experiment II only speaks to the acoustical processing hypothesis by comparison with other experiments; some further investigation of the role of purely acoustical features is desirable (see Haggard and Parkinson 1971), but it is hard to see what purely acoustical features could have been introduced by this experiment.

Finding no dependence of REA upon voice of the preceding trial argues against the critical innovation of Experiment II having been an increase in effective channel separation or some other general prerequisite of REAs, but the possibility deserves closer scrutiny. Failure to obtain greater REA on trials following trials in the same voice does not rule out all ways in which this sort of factor might operate; the fact of a pitch difference between the two ears and hence the two voices

might have over-ruled the effects of alternation or sequence of voices upon attention. Thus we wish to know whether effective channel separation is at all relevant in this experiment. If it were we could expect the REA to have been largely created by special errors in response to left ear stimuli—intrusions of what was actually presented to the right ear.

An analysis was made of the types of error in the experiment, and the intrusions expressed as a proportion of errors made when listening to each ear, for each of the 40 subjects. Intrusions did constitute a significantly preferred class of error; of the 80 scores only 13 produced a proportion of intrusion less than the chance expectation of 1 in 5 (two ties). Of the 40 subjects 20 produced a greater proportion of intrusions while listening to the right ear, 19 while listening to the left (1 tie). These data were obtained from approximately 50 errors in each case, thus the REA does not result from intrusions.

Intrusions might be related to both voice and sequential condition, so the analysis was broken down further. There was a slight overall advantage for the words in the high voice (2%) but this did not interact with the other effects so it was not included as a factor in the analysis of the results. However, a check was made to see whether this advantage led to more intrusions. For 19 subjects the high voice on the unattended ear produced relatively more intrusions than the low; for 19 the reverse was true (two ties), so the difference between voices can be discounted.

There was a significant dependance of intrusions upon the sequential condition. Trials following trials in the same voice produced fewer intrusion errors (6 exceptions, two ties) but they did not produce significantly greater REA on purely confusion errors (13 exceptions to REA, 10 ties) than the trials following a trial in the different voice (16 exceptions to REA, 7 ties). Hence we can say that, while attention and intrusion are operating in the experiment and they do depend upon the sequential condition, any relationship any of these factors might have to the REA is very slight and unlikely to explain the emergence of REA in the present experiment.

We may now consider the implications of the finding that only a trigger feature signalling *potential* encoding is required to produce REA. We will not be concerned with such general prerequisites as interaural competition. Two possible ways of explaining the finding would be as follows. A selector might direct stimuli containing the trigger feature to a centre able to decode them, putatively in the left hemisphere. A more "passive" and hence preferable interpretation would be that in identification of the stimulus the analysers for encoded features would have a high weighting; any input containing a trigger feature will produce some output from the decoding analyser. If this analyser has a strong left hemisphere dominance and the decision task employs its output an REA will be obtained. The irrelevance of the actual encoding rules out any explanation in terms of fully independent parallel processing by the two hemispheres or of some non-specific superiority of the left. Thus the right hemisphere appears not to process the encoded features at all rather than to do them less accurately or slower.

Comparison with Shankweiler and Studdert-Kennedy's experiment suggests that the trigger feature does not assure some overall dominance of one hemisphere

in processing for the whole stimulus word. Analogously experiments by Darwin (1971) show that formant transitions do not guarantee REA for the voicing distinction, although other features may. Thus the REA obtained in the present type of experiment applies to functional asymmetries at the level of the extraction of the linguistically significant features in terms of which the phonemes are specified. Tasks emphasizing the higher processes as well might produce larger REAs but as phonemic distinctions are defined as minimal carriers of meaning differences we would never claim that the asymmetry operates "below the level of meaning". Any phonologically admissible sequence is *potentially* meaningful for a device concerned with the first-level decoding of speech.

Our earlier suggestion that dichotic experiments would inform us about interhemispheric relations rather than speech processing may have been premature.

In conclusion we can say that we have ruled out the possibility that the REA for speech signals is simply a function of the linguistic status of the items used in the perceptual test. Simple acoustical properties appear not to be involved. Encoding seems to be a necessary stimulus property; adding an element of encoding to vowels produces an REA which does not depend upon any more general factor such as effective channel separation. A genuine need for perceptual decoding in the long or short term does not appear to be necessary, arguing in favour of a preferential processing in one hemisphere of encoded aspects of stimuli marked by trigger features. The restriction of the REA to tasks where the "trigger features" are relevant suggests that the lateralisation can occur at a low level in the perceptual system before the organisation of a response.

I would like to thank Professor O. L. Zangwill for research facilities, JSRU-GPO/CESG, and Corpus Christi College, Cambridge for research support, and Dr F. S. Cooper of Haskins Laboratories for hospitality and facilities.

Reference

BROADBENT, D. E. (1962). Attention and the Perception of Speech. *Scientific American* **206,** No. 4, 143–51.

DARWIN, C. J. (1969). Auditory perception and central dominance. Unpublished Ph.D. Thesis University of Cambridge.

DARWIN, C. J. (1971). Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology* **23,** 46–62.

HAGGARD, M. P. (1966). Perceptual selectivity and the recognition of speech. Ph.D. Thesis, Cambridge.

HAGGARD, M. P. and MATTINGLY, I. G. (1968). A simple program for synthezising British English. *Institute of Electrical and Electron Engineerings Audio* 16, 95–9.

HAGGARD, M. P. and PARKINSON, A. M. (1971). Stimulus and task factors in the perceptual lateralization of speech signals. *Quarterly Journal of Experimental Psychology,* **23** (in Press).

HALWES, T. (1969). Effects of dichotic fusion in the perception of speech. Unpublished Ph. D. Thesis, University of Minnesota.

KIMURA, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology,* **15,** 166–71.

KIRSTEIN, E. and SHANKWEILER, D. P. (1969). Selective listening for dichotically presented consonants and vowels. Paper to E. P. A., Philadelphia, April 1969.

LADEFOGED, P. and BROADBENT, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America,* **29,** 98–104

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P. and STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431–61.
LINDBLOM, B. E. F. and STUDDERT-KENNEDY, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, **42**, 830–43.
MACNEILAGE, P. (1970). Motor control of the serial ordering of speech. *Psychological Review*, **77**, 182–96.
SHANKWEILER, D. P. and STUDDERT-KENNEDY, M. (1967). Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, **19**, 59–63.
SHANKWEILER, D. P. and STUDDERT-KENNEDY, M. (1970). Hemispheric specialisation for speech perception. *Journal of the Acoustical Society of America*, **48**, 579–94.

Received 25 *August* 1970

ANNOUNCEMENTS

Conference on "Inhibition and Learning"

The Laboratory of Experimental Psychology, University of Sussex, and the Experimental Analysis of Behaviour Group are sponsoring a 3 day conference at Sussex from 5–7 April 1971. Topics will include the role of inhibition and related physiological mechanisms in discrimination learning, operant conditioning and classical conditioning.

Provisional speakers include M. E. Bitterman, R. J. Douglas, E. Grastyan, E. Hearst, R. L. Isaacson, J. Konorski, N. J. Mackintosh, R. A. Rescorla, H. S. Terrace and A. Wagner.

For further information please write to: *Dr M. S. Halliday, Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex BN*1 9*QY*.

5th International Congress of the International Ergonomics Association

The Fifth International Congress of the International Ergonomics Association (I.E.A.), organized by the Nederlandse Vereniging voor Ergonomie, will be held on 4–8 June 1973, at the International Congrescentrum RAI, in Amsterdam, The Netherlands. Congress Secretariat: Organisatie Bureau Amsterdam NV, Postbus 7205, Amsterdam.

Quarterly Journal of Experimental Psychology (1971) **23**, 46–62

EAR DIFFERENCES IN THE RECALL OF FRICATIVES AND VOWELS

C. J. DARWIN

Department of Psychology, University of Connecticut and Haskins Laboratories

Two experiments on the free recall of dichotically presented synthetic speech sounds are reported. The first shows that the right ear advantage for initial fricative consonants is not simply a function of the recognition response class, but that it is also a function of the particular acoustic cues used to achieve that response. This is true both for the whole response, and for the constituent phonetic features. The second experiment shows that when both the response class and the particular stimuli presented on certain trials are held constant, the right ear advantage for the constant stimuli can be influenced by the range of other stimuli occurring in the experiment. Vowels show a right ear advantage when, within the experiment, there is uncertainty as to vocal tract size, but they show no ear advantage when all the vowels in the experiment are from the same vocal tract. These results are interpreted as demonstrating that there are differences between the ears, and probably between the hemispheres, at some stage between the acoustic analysis of the signal and its identification as a phonetic category.

Introduction

Under certain conditions, sounds which enter one ear may subsequently be more efficiently recalled or recognized than similar sounds entering the other ear (Kimura, 1961*a*,*b*, 1964). Differences between the ears tend to be obtained more reliably when different sounds enter the two ears simultaneously than when only one ear is stimulated, either with one (Corsi, 1967) or with two simultaneous signals (Shankweiler, in press). Monaural stimulation *can* give significant ear differences but such experiments have required larger numbers of subjects than the usual dichotic paradigm (Bakker, 1968, 1970).

The type of stimulus material used is probably the only determinant of *which* ear gives better performance. In similar recognition paradigms the right ear does better for digit triads (Broadbent and Gregory, 1964) while the left does better for orchestrated melodies (Kimura, 1964) and simple pitch patterns (Darwin, 1969). In free recall, the right ear again does better for digit triads (Kimura, 1961*b*) while the left does better for familiar melodies (Kimura, 1967) and simple pitch sweeps, whether carried on a word or on a non-verbal timbre (Darwin, 1969).

Since patients with vocal speech impaired when their *right* hemispheres are anaesthetized show an advantage for the *left* ear in free recall of digit triads (Kimura, 1961*a*), some link between the ear difference effect and cerebral dominance must be assumed. Authors differ on the nature of this link. Some attribute it to perception (Kimura, 1961*b*), some to short-term memory (Inglis, 1962), others to attention (Treisman and Geffen, 1968). Some authors have implicitly denied the stimulus specificity of the direction of the effect, and claim that there is a general

tendency to report material entering the right ear before that entering the left (Oxbury, Oxbury and Gardiner, 1967).

One important limitation of free recall experiments was pointed out by Inglis (1962). Serial-order effects (see e.g. Broadbent, 1958) could account for ear differences in a free recall paradigm if there were some tendency to report certain types of material from a particular ear first. Bryden (1963) controlled for serial order effects and found a smaller, though still significant, residual advantage for the right ear with digit sequences. Thus, while serial order effects account for some of the ear difference in a free recall paradigm, they neither explain why the sounds from one ear are recalled first, nor why there is a residual difference. The tendency to report one ear first could derive from whatever causes this residual ear difference.

This residual effect may be due to differences in the efficiency with which material is either perceived or remembered. Making the distinction between perception and memory in terms of the first and second ear reported, Bryden (1967) summarizes the available data, and shows that there is no evidence that the ear difference effect is any smaller on the first than on the second reported ear. Darwin (1969) also failed to find any such evidence both for material recalled better from the right and from the left ears.

Treisman and Geffen (1968) suggested that the ear difference effect arises because of an unequal distribution of attention, the left hemisphere finding it easier to attend to the right ear than the left ear. If this were so we would expect sounds which are more easily separated by selective attention to show a greater ear difference than those which are more difficult to separate. Kirstein and Shankweiler (1969), however, find that when a subject is asked to report the sounds from a particular ear, he makes fewer errors of attention for vowels than for consonants, but consonants show a greater right ear advantage than vowels. Selective attention may interact with the mechanisms responsible for the ear difference effect but it is not a basic cause.

Kimura's (1961*b*) explanation of the ear difference effect as reflecting differences in the efficiency with which material is perceived (in the sense used above) in the two hemispheres can account for all the available data that has been obtained with adequate experimental procedures, provided that we make the assumption that the experimental differences demonstrated between the ears can be attributed to differences between the two hemispheres. No alternative explanation can do so well. What then is the nature of this "perceptual" difference? At what stage in the varied processes of perception do differences between the ears and between the hemispheres appear?

The right ear advantage does not depend on the material being meaningful. Significantly greater scores for the right ear than for the left have been detected in free recall paradigms for initial and final stop consonants (Shankweiler and Studdert-Kennedy, 1967*a*,*b*) and for laterals and semi-vowels (Haggard, 1969) in a simple nonsense syllable context. The right ear advantage for stops remains when order of report is controlled by a suitable method of scoring (Darwin, 1969), or by pre-instructing order of report (Kirstein and Shankweiler, 1969). However, these experiments do not tell us whether the difference between the ears occurs

before or after the sound has been categorized as a particular phoneme. The failure of vowels to give a right ear advantage in free recall (Shankweiler and Studdert-Kennedy, 1967*a*; Darwin, 1969) is not relevant here since vowels differ from consonants in both their acoustic structure and their phonological class. Vowels and consonants could have different ear asymmetries at some level *after* they have been classified as phonemes. This paper examines whether there are differences between the ears in some perceptual process which occurs *before* classification of a sound as a phoneme.

Analytically, the sounds of speech form a sub-set of the sounds of the environment since they are subject to phonetic constraints deriving from the anatomy and physiology of the vocal tract and to phonological and allophonic constraints imposed by particular languages. Maximum efficiency in perception will only be obtained if these constraints are utilized. However, to preserve the efficient perception of sounds not subject to these constraints, some functional division is required in the perceptual system so that one part may deal with the special problems of speech whilst the other remains free to deal with the remaining sounds.

The phonetic constraints are of two main types, both of which lead to a complex relationship between the perceived phoneme and the acoustic signal. In one case a complex relation arises because the articulatory specifications for some phonemes are incomplete (for example, only a general movement of the lips and jaw is specified for bilabial stops); the articulators which are not specified can then assume a wide variety of positions with a correspondingly wide variety of acoustic sequelae. In the second case the complex relation arises from the variation in size and shape of the vocal tracts producing the sound. The first set of relations has been extensively studied, and the word "encoded" has been used (Liberman *et al.*, 1967) to describe this particular lack of acoustic invariance. The second type of variability has received relatively little study. However, the relationship is not likely to be a simple one since, for example, women's vocal tracts are not only smaller than men's, but have different relative proportions (Chiba and Kajiyama, 1941). So when a vowel is spoken by two different individuals with the same articulatory gestures, the formant frequencies for one cannot, in general, be obtained by multiplying each formant frequency of the other's by a constant multiple. This multiple varies between speakers, between vowels and between individual formants (Mattingly, 1966; Fant, 1966). The perceptual system at least partially compensates for these perturbations, since it can accommodate some independent variation in the range of the first two formants (Ladefoged and Broadbent, 1957).

These are by no means the only problems for the speech recognition system, but as they are specific to speech they offer the opportunity of separating speech and non-speech perceptual mechanisms, and asking whether they are equally the prerogative of the two ears and of the two hemispheres. The first experiment asks whether the ear advantage is the same for sounds perceived as the same phoneme, but requiring to different extents Liberman's "decoder", while the second experiment asks the same question of vowel sounds from different sized vocal tracts.

Experiment I. Fricatives

Fricatives are well suited to the purpose of this experiment since there are two main cues which contribute to their perception. The first, and perceptually most significant is the spectral peak of the friction itself (Harris, 1958; Heinz and Stevens, 1961); this peak shows relatively little variation with vowel context. A secondary cue is the formant transitions to adjacent vowels. These show much more contextual variation with vowel context as they depend on the shape of the whole vocal tract. In both voiced and unvoiced fricatives they only assume a major role in distinguishing /f, v/ from θ, ð/, although they do contribute to the intelligibility of the other distinctions. Fricatives synthesized with appropriate formant transitions are generally more intelligible than those synthesized without, although the latter are still highly intelligible provided that the /f,v/-/θ, ð/ distinction is not required.

Liberman *et al.* (1967) hypothesized that only those aspects of speech which show appreciable contextual variation give a right ear advantage. This predicts that fricatives containing the appropriate formant transitions will show a right ear advantage, while those without such transitions will not.

Method

The experimental tape was prepared on the Haskins parellel formant synthesizer. The six fricatives /f, s, ʃ, v, z, ʒ/ were used in the syllabic frame /-ɛp/. /θ/ and /ð/ were not used because they are highly confusable with /f/ and /v/, respectively. There were four stimulus conditions:

(1) with appropriate friction and appropriate formant transitions;
(2) as (1) but with an instantaneous transition into the vowel, which was extended to occupy the time previously allocated to the transition;
(3) as (2) but with the vowel deleted, leaving only the steady-state friction;
(4) as (1) but without the friction, leaving formant transitions and vowel. This condition sounded like plosives rather than fricatives.

The steady-state friction lasted 45 msec, the transitions 30 msec, and the final syllable 120 msec.

The sounds were assembled into a dichotic tape, using a computer program (Mattingly, 1968) which first laid down marker pulses on the recording tape, and then synthesized utterances in a predetermined sequence as the marker pulses were detected. This method allows individual dichotic pairs to be aligned almost perfectly in time, while the use of synthetic speech allows accurate control of the amplitudes and duration of the sounds.

Each sound was paired twice with every *other* sound in its own stimulus condition to give a basic experimental tape of 240 trials, the second half of which was the same as the first, but with the trial order reversed. This whole experimental tape was taken by each subject twice. Prior to the main experiment the subjects were practiced in identifying the sounds with the following letters: f, s, sh, v, z, j, p, b, d. A pilot experiment showed that the letters p, b, d were most readily assigned to the quasiplosives which constituted condition (4). This condition was not basic to the purpose of the experiment, but was included in case none of the fricative conditions gave a significant ear advantage. When the subjects were scoring above 75% on these single sounds they were given 10 practice trials with dichotic pairs. They were told to write down the two sounds they heard putting their more confident choice first. They could write down the same response twice if they wished. They were asked to try to maintain a neutral attention before each trial, rather than listen for only one ear. After the ten practice trials, if they had no questions and had not obviously flouted the instructions, they went on to the main test trials, which came in

16 blocks of 30 trials. Half the subjects started with the headphones reversed, and all subjects reversed their headphones after every 4 blocks.

The experiment was taken by 1 left-handed and 13 right-handed undergraduate and graduate subjects. No subject had any hearing defect to the best of his knowledge, and none had a difference of more than 5 dB between the ears for the threshold at 1500 Hz measured by the method of limits.

Results

Statistical tests are taken from Siegel (1956) and are all two-tailed. Unless otherwise stated, the test used is a Wilcoxon T-test for matched pairs. The overall percentages correct for the first and second responses together are given in Table I.

TABLE I

Percentages correct for total scores on both responses by stimulus condition

	Stimulus condition			
Ear	1 Friction transition vowel	2 Friction vowel	3 Friction	4 Transition vowel
Left	47·4	44·4	45·4	58·6
Right	53·2	46·6	46·6	63·6
Right − left	5·8	2·2	1·2	5·0
Right + left	50·3	45·5	46·0	61·1

A Friedman analysis of variance on total right minus left ear scores between the four stimulus conditions is significant ($P < 0·01$). The total score on the right ear is significantly higher than that on the left for condition (1) ($P < 0·01$), and condition (4) ($P < 0·05$), but not for either condition (2) or (3) ($P > 0·1$). This picture holds both with and without the left-handed subject. Condition (1) gives a significantly greater right ear advantage than either condition (2) ($P < 0·02$) or condition (3) ($P < 0·1$). Adding formant transitions thus increases the score more on the right ear than on the left. The left-handed subject shows a large effect in the opposite direction with conditions (2) and (3) showing a greater right ear advantage than condition (1). He is omitted from all remaining statistics.

The total scores show a very significant tendency for the right ear to score higher on condition (1) than on condition (2) ($P < 0·001$), but only a slight tendency for the left ear to do so ($0·1 > P > 0·05$). A similar pattern prevailed between conditions (1) and (3) but not between conditions (2) and (3). Performance on the right ear is significantly better when formant transitions are added, whilst that on the left ear is not. Thus only the right ear can utilize the additional information present in the formant transitions effectively.

Since the preceding analysis has been made in terms of simple percentage correct scores the differences found between the various stimulus conditions may be due partly to changes in preferred order of report, although it is difficult to think of any interesting reason why this should be so. To counter this objection,

however, a scoring system was devised which compensated for order of report effects. These "D-scores" are described in the Appendix. D_1 scores reflect the first channel reported and D_2 the second. A positive D-score indicates a right ear advantage.

TABLE II

Mean D *scores for fricatives by stimulus condition*

	Stimulus condition		
	1	2	3
	Friction transition vowel	Friction vowel	Friction
D_1	−0·253	−0·050	0·109
D_2	0·161	0·072	0·022
$D_2 - D_1$	0·092	0·122	0·087

Positive D score indicates right ear advantage, subscript denotes order of report.

A Friedman analysis of variance on the D_1 scores is almost significant ($0{\cdot}1 > P > 0{\cdot}05$), but fails significance on the D_2 scores ($P > 0{\cdot}1$). The significance level of individual Wilcoxon T-tests on these scores is therefore not reliable. The following significance levels are given, however, as an indication of the pattern of the results. The important differences, those between condition (1) and conditions (2) and (3), respectively, appear large and show apparent significance levels of less than 0·025 for the D_1 scores. As in the percentage correct analysis there is a large difference between conditions (1) and (2) for the right ear scores ($P < 0{\cdot}002$), but a small one for the left ear scores ($P > 0{\cdot}1$).

Although the D scores are too variable to allow these significance levels to be accepted, the overall pattern of results is almost identical to that of the percentage correct scores. Since the D scores compensate for order of report effects, it is unlikely that the significant patterns seen in the percentage correct scores are attributable to a change in order of report preferences. It seems more probable that the D scores are inherently more variable than the simple percentage correct from which they are derived.

In summary, a similar pattern of results is obtained with both simple percentage correct scores and a more complicated score which makes some compensation for the order in which the two ears are reported and the overall level of performance. The right ear advantage is greater when appropriate formant transitions are present than when they are absent. The presence of a succeeding vowel in the absence of formant transitions, however, does not appear to influence the ear advantage. The ear difference effect is thus not simply a function of the recognition response class, but is also influenced by the particular cues used to achieve a given response. Moreover, the results are as predicted by Liberman *et al.*'s encoding hypothesis in that only those sounds with formant transitions show a right ear advantage.

So far in this analysis we have taken as correct a response which has both the appropriate voicing and place of articulation. It is of some interest to see whether there are ear advantages for these two dimensions independently. There is convincing psychological evidence that the traditional phonetic feature system is implicated in processes of perception (Miller and Nicely, 1955) and short-term memory (Wickelgren, 1966). If the ear difference indeed reflects differences in the perceptual efficacy of the two ears, these differences may be present not only for the perception of the phoneme as a whole, but also for the perception of its constituent features.

In a dichotic listening experiment using stop consonants, Halwes (1969) found that a large proportion of errors arose from a failure to combine features correctly rather than from a failure to extract them. Many "incorrect" responses in Halwes' experiment consisted of a feature from one ear combined with a feature from the other ear. Perhaps when, as in this fricatives experiment, a correct response is scored only when both voicing and place of articulation are correct, the ear difference is due to a difference in the efficiency with which the two features are combined into a response rather than to any differences in the efficiency with which they are actually extracted. If this were entirely the case we would expect there to be no residual ear difference if the ear effects for the two dimensions are assessed separately. On the other hand, it is possible that there are differences between the ears in the efficiency with which the features are actually extracted, in which case we *would* expect ear differences when we analyse the features separately.

The results of the fricatives experiment were accordingly scored to provide separate analyses of the voicing and place of articulation dimensions. The dimension not under consideration was made irrelevant both in the stimulus and in the response. This procedure is necessary if the analyses of the two dimensions are to be truly independent.

Analysis of place of articulation was carried out in terms of overall percentage correct, making voicing irrelevant in both the stimulus and the response. A Friedman analysis of variance gave a significant overall variation over stimulus conditions for right minus left ear percentage correct scores ($X_r^2 = 7{\cdot}55$, $df = 2$, $P < 0{\cdot}05$). As in the main analysis the only condition to show a significant right ear advantage was the first, that which had friction and formant transitions ($T = 4\frac{1}{2}$, $n = 13$, $P < 0{\cdot}005$). Neither group 2 nor group 3 showed a significant right ear advantage ($P > 0{\cdot}1$). There was a significant difference between the first group and the average of the other two in this respect ($T = 14\frac{1}{2}$, $n = 13$, $P < 0{\cdot}05$). Analysis in terms of D scores was not made because of the large variance with only three response alternatives.

For the voicing dimension the only trials which contribute differentially to the ear difference are those on which the two stimuli have different voicing, but the two responses have the same voicing. Only one of the stimuli has then been incorporated into the response. A Friedman analysis of variance on the difference between right and left ear incorporations of voicing for the three fricative conditions is significant ($X_r^2 = 7{\cdot}0$, $df = 2$, $P < 0{\cdot}05$). Individual T-tests show that voicing is incorporated more often from the right ear than from the left in both the

first ($T = 13\frac{1}{2}$, $n = 12$, $P < 0.05$) and the second ($T = 11$, $n = 12$, $P < 0.05$) stimulus conditions (the two with the succeeding vowel). There is no significant right ear advantage for the third condition with the isolated friction ($T = 20\frac{1}{2}$, $n = 11$, $P > 0.1$). There is a significant difference between groups 2 and 3 in this respect ($T = 12$, $n = 13$, $P < 0.02$) but not between any of the others. Combining the first two groups gives a highly significant advantage for the right ear ($T = 1\frac{1}{2}$, $n = 12$, $P < 0.002$) and a significant difference between their mean and the third group ($T = 10$, $n = 11$, $P < 0.05$). Thus the voicing dimension is reported more accurately from the right than from the left ear only when there is a succeeding vowel.

For fricatives there is thus a dissociation between the stimulus conditions necessary to give a right ear preference for place of articulation and those necessary to give one for voicing. Formant transitions are necessary for the former but a succeeding vowel suffices for the latter. However, these conclusions must be qualified by their possible contamination with changes in order of report preferences since they are based on an analysis of percent correct scores.

Discussion

The main result of this experiment is that the right ear advantage is not determined solely by the recognition response, but is also influenced by the particular sound used to achieve that response. This appears to be true both for the phonetic response as a whole and for the individual articulatory features which constitute that response. Moreover, the particular acoustic signals which must be present for voicing or for place of articulation to show a right ear advantage are different. For place of articulation appropriate formant transitions must be present, whilst for voicing a succeeding vowel suffices. This dissociation suggests that the difference between the ears is occurring before or during the classification of the sound into features, and that it is not simply a consequence of an overall ear difference for the phonemic response. In particular the presence of a right ear advantage for voicing under condition 2, when there is no overall right ear advantage for the entire phoneme argues that the ear difference for the individual features is not a consequence of the ear advantage for the entire response, but rather that the ear advantage for particular features logically precedes that for the entire response.

If differences between the *ears* are not simply a function of response class, can the same be said of differences between the *hemispheres*? Unfortunately, no. An important assumption in the interpretation of ear differences is that there is a functional decussation of the auditory pathways. Although there is electrophysiological evidence which shows a statistical decussation in sub-human species both for evoked potentials (Tunturi, 1946; Rozensweig, 1951) and for single unit recording (Hall and Goldstein, 1968), the main evidence we have that this decussation is both present in man and sufficient to reveal inter-hemispheric differences is the results of dichotic listening experiments. The most convincing demonstration occurs in patients with a section of the corpus callosum. These patients can report verbal material equally well from either ear when only one ear is stimulated at a time, but can report practically nothing from the left ear

when similar verbal material is played simultaneously into both ears (Milner, Taylor and Sperry, 1968). Moreover, this weakening of the left ear response is dependent on the nature of the sounds in the other ear. As the sounds in the right ear are progressively distorted, performance on the left ear improves (Sparks and Geschwind, 1968).

Normal subjects show much smaller ear differences than the commissurectomized patients when undistorted digit sequences are played in both ears (Milner *et al.*, 1968; Kimura, 1961*b*). Normal subjects also show an ear difference effect which is dependent on the nature of the competing stimulus. Initial and final plosive consonants give a reliable right ear advantage when they are opposed by another such consonant (Shankweiler and Studdert-Kennedy, 1967*b*); however, plosive consonants embedded in a nonsense word and opposed by white noise give no ear difference (Corsi, 1967). An unpublished experiment by the present author showed no ear difference between the ears using initial plosives rather than embedded ones, with noise on the other ear. Thus the ear difference effect *is* influenced by the nature of the competing stimulus.

The simplest explanation of these effects is that in normal subjects considerable information about the sounds on the left ear can be transmitted across the commissures to the left hemisphere. The commissurectomized patients, being deprived of this path, must rely entirely on the direct ipsilateral path. The efficiency of this latter path is critically dependent on the nature of the sounds on the two ears. With no sound on one ear, it can function well, but as progressively less distorted speech is introduced on the other ear, it becomes less and less efficient.

A significant difference between scores from the two ears can be interpreted as showing that there is some difference between the hemispheres, and that the sounds on each ear have gone predominantly to their opposite hemispheres. However, if there is no significant difference between the ears, we cannot attribute this failure with any confidence to either an equivalence of the two hemispheres or to a failure of the relevant pathways to decussate sufficiently to reveal an inter-hemispheric difference. The differences in ear advantage between the various stimulus groups reported in this experiment could then be due either to a difference in the degree to which the two hemispheres are implicated in their processing, or to a difference in their abilities to produce a functional decussation of the relevant pathways. We can only conclude that the former is true, and thus that the hemispheres differ in their ability to classify phonemes if we have independent evidence that those sounds which did not give a right ear advantage were *in principle* capable of revealing any inter-hemispheric difference that there might have been.

All the sounds which failed to give an ear advantage for a particular feature in this experiment had a steady state along the physical dimension relevant to that phonetic feature. Thus place of articulation only shows an ear advantage when it is cued by a moving pattern of formant transitions, while the voicing feature only shows an ear advantage when it is cued by a sound which may be only partially voiced. Perhaps no steady-state discrimination can give an ear difference. The absence of any ear difference for steady-state vowels, whether in CVC context or in isolation (Shankweiler and Studdert-Kennedy, 1967*a*,*b*), and of very brief

duration (Darwin, 1969) supports this idea. Furthermore, Darwin (1969) found only tenuous evidence for a left ear advantage for recall of steady-state non-verbal timbres similar to those whose discrimination was more impaired after right, than left temporal lobectomy (Milner, 1962). If an ear advantage can be demonstrated for steady-state sounds, we will have more justification for assuming that the steady-state sounds used in this fricatives experiment were in principle capable of showing ear differences.

We must now face the logical difficulty that without further assumptions we cannot tell whether any change made in the stimulus conditions which produces an ear advantage is having its effect through changing the conditions necessary to reveal differences between the hemispheres, or through changing the nature of the task in such a way as to implicate mechanisms for which the hemispheres do in fact differ.

One reasonable assumption is that the functional decussation of the auditory pathways is determined only by the particular sounds which are presented on any one trial, and is not influenced by the range of sounds which may occur in the experiment. In other words, if we know from the fact that they give an ear advantage that there is good decussation for a particular dichotic pair of sounds in one experiment, we can remove some of the other dichotic pairs from the experiment without changing the functional decussation for that particular pair. In contrast, the number of different dichotic pairs used in an experiment will generally alter the complexity of the task, and so perhaps alter the relative contribution of either hemisphere. If, then, we can show that greater ear advantages can be obtained for some sounds when the number of different stimuli used in the experiment is changed, we might assume we are measuring a change in inter-hemispheric ability rather than a change in the functional decussation of the auditory pathway.

If, then, the steady-state sounds used in this, and other experiments, have failed to show any ear difference solely because of inadequate functional auditory decussation, we should not expect such sounds to show an advantage when only the complexity of the perceptual discrimination is changed. The next experiment attempts to demonstrate that the ear advantage *is* influenced by the complexity of the perceptual discrimination by changing the range of vocal tract sizes that a set of vowels can come from.

Experiment II. Vowels from Different Sized Vocal Tracts

There is a rough correlation between voice pitch and formant frequencies, since women and children have higher voices and smaller vocal tracts than men. This correlation is utilized in estimating vocal tract size (Fujisaki and Kawashima, 1969). A recent experiment by Haggard (1971a) shows that when vowel perception depends on the fundamental frequency of the vowel, there is a right ear advantage under free recall conditions. Steady-state sounds are here showing a right ear advantage, when there is a difference in pitch between the two ears. Unfortunately for the present argument, this difference in pitch is a reasonable candidate for a factor which changes the conditions necessary to reveal the ear difference effect, as well as one which alters the perceptual complexity of the task. Can we show a right

ear advantage for steady-state vowels which have the same pitch on either ear? The most direct way to answer this question is to use sets of vowels from two different sized vocal tracts.

Method

The five vowels /i, ɛ, æ, a, ʌ/ in the context /ən-t/ were synthesized on the Haskins parallel formant synthesizer using only the first two formants. Two sets of these five words were made, the formant frequencies for one set being 25% higher than those for the other set. The formant values are given in Table III.

Two different experimental tapes were then constructed. On one tape, each sound was paired with every other sound except itself and its phonemic homologue from the other vocal tract. On the other tape, only the sounds from the smaller vocal tract were used, and each sound was paired with every other sound except itself. The first tape had 160 trials and the second 40. The order of the trials on the second tape was exactly the same as the order of those trials on the first tape in which both sounds came from the smaller vocal tract.

The first tape was taken twice by one group of 18 subjects, and the second tape was taken twice by a second group of 18 subjects. All subjects were right-handed, native speakers of American English, who to the best of their knowledge had no hearing defects. The instructions and training they received were similar to those used in the fricatives experiment. The words used to identify the sounds were "a nit, a net, a gnat, a knot, a nut", and both groups of subjects used the five letters "i, e, a, o, u" as their responses. Those who took the first tape had training in identifying the sounds from both vocal tracts, whereas the second group of subjects were only introduced to the sounds from the smaller vocal tract. The usual counter-balancing procedures were observed.

TABLE III

Formant frequencies for vowels in Experiment II

Vowel	Large vocal tract F1	Large vocal tract F2	Small vocal tract F1	Small vocal tract F2
/i/	386	2078	489	2540
/ɛ/	537	1845	666	2307
/æ/	666	1695	844	2156
/a/	718	1075	894	1312
/ʌ/	640	1232	794	1541

Results

Five stimulus conditions are distinguished in the results. Four come from the first group of subjects and correspond to whether the dichotic pair had sounds from (1) the larger vocal tract only; (2) the smaller vocal tract only; (3) the larger on the left ear and the smaller on the right; (4) the smaller on the left and the larger on the right. The fifth condition corresponds to the second group of subjects who had the smaller vocal tract on both ears all the time. The overall percentage correct and the D scores are given in Tables IV and V, respectively.

The overall superiority for the right ear for the first group of subjects (summing over the first four stimulus conditions) is significant both on percentage correct ($P < 0.001$) and on D_1 scores ($P < 0.01$). For the second group of subjects there is no significant right ear advantage on either score ($P > 0.1$).

A Friedman analysis of variance over the first four stimulus conditions is significant for differences in percentage correct ($P < 0.01$) and D_1 scores ($P < 0.02$). The variation in overall level of performance, however, is barely significant ($P < 0.1$). Individual Wilcoxon T-tests show that ear differences are significant for the first and second stimulus conditions separately on overall percentages correct ($P < 0.01$) and on D_1 ($P < 0.05$ and < 0.06, respectively).

For conditions (3) and (4) combined, when the two-ears had different vocal tracts, the right ear did significantly better than the left ($P < 0.002$) but there was also a significant tendency for the vowels from the small vocal tract to be recalled

TABLE IV

Overall percentages correct in Experiment II by dichotic pair composition

Vocal tract size on Left ear	Right ear		Overall percentage correct on Left ear	Right ear	Right-left	$P(L=R)$
Large	Large	(1)	46·7	50·9	4·2	<0·01
Small	Small	(2)	45·8	50·4	4·5	<0·01
Large	Small	(3)	45·7	56·2	10·6	<0·002 (3)–(4)
Small	Large	(4)	54·0	51·2	−2·8	
Total			48·1	52·2	4·1	<0·001
Small	Small	(5)	54·0	53·4	−0·6	>0·1

TABLE V

D *scores for Experiment II by dichotic pair composition*

Vocal tract size on Left ear	Right ear		D_1	D_2	$p(D_1 = 0)$	
Large	Large	(1)	0·083	0·078	<0·06	<0·002 (1)–(2)
Small	Small	(2)	0·103	0·067	<0·05	
Large	Small	(3)	0·226	0·206		<0·002 (3)–(4)
Small	Large	(4)	−0·062	−0·089		
Small	Small	(5)	−0·060	−0·013	>0·10	

better than those from the larger ($P < 0.05$). This difference is not present when the two ears receive vowels from the same vocal tract as in conditions (1) and (2). It is not then due to markedly poorer intelligiblity for the smaller vocal tract.

There is a significantly greater right ear advantage for the vowels in condition (2) than in condition (5) both for percentages correct ($P < 0.05$) and for D_1 scores ($P < 0.02$) on Mann-Whitney U-tests. But there is no difference between the averages of conditions (1) and (2) vs. conditions (3) and (4) ($P > 0.1$). In other words, the right ear advantage for vowels in this experiment depends on the nature of the discrimination within the framework of the whole experiment rather than within the individual trial.

A reliable right ear advantage for steady-state vowels, therefore, can be obtained

when there is uncertainty within the experiment as to what size vocal tract has produced them. But this right ear advantage is not influenced by whether on a particular trial the two alternative sizes of vocal tract are in fact present.

Discussion

This experiment confirms Haggard's finding that vowels can give a right ear advantage. Whether the advantage appears or not in this experiment depends on the complexity of the perceptual discrimination, rather than on the particular sounds used on any one trial. On the assumption that the sounds used for the second group of subjects were in principle capable of showing a right ear advantage, we can conclude that the hemispheres do differ in their ability to classify vowels from different sized vocal tracts. This assumption seems reasonable, since identical sounds did give a right ear advantage when played to the first group of subjects, as part of a larger experiment.

The assumption that was necessary in interpreting the results of the fricatives experiment in terms of differences between the two hemispheres has received some justification, since the vowels used here are cued mainly by a steady-state. More direct confirmation of this could perhaps be obtained by using steady-state friction from different sized vocal tracts.

Can we draw any conclusions about the stage or stages in perception at which ear or hemisphere differences become apparent? The ear difference effect is not solely a function either of the stimulus, or of the response, but rather of the processes which must mediate between the two. The fricatives experiment showed that it did not depend on the response category alone, since whether or not it appeared either for the entire phonetic response or for one of the constituent dimensions of voicing and place of articulation depended on the presence of particular acoustic cues. The vowel experiment described here shows that the effect does not depend solely on either the stimuli presented on a particular trial, or on the response category, since the same stimuli do or do not show a right ear advantage depending on the complexity of the relationship between the stimuli and the responses.

A similar conclusion has been reached by Studdert-Kennedy and Shankweiler (1970) on the basis of a feature analysis of a dichotic experiment with stop consonants. They, with Halwes (1969), find that a large proportion of errors arise from inappropriate combination of correctly extracted features. They suggest that this arises because acoustic features can be extracted correctly in either hemisphere, but that they can only be related to phonemic features and assembled into a phonemic response in the left hemisphere.

More direct evidence that particular acoustic features themselves are not entirely responsible for the ear difference effect comes from an experiment by Haggard (1971b). Haggard shows that when the voicing dimension is cued only by a change in pitch (Haggard, Ambler and Callow, 1970) in a dichotic listening paradigm, the recall of this feature shows a right ear advantage. Since Darwin (1969) has shown that simple pitch sweeps give a *left* ear advantage when carried on a word, but do not cue a phonemic distinction, it seems likely that the pitch sweeps which cued voicing in Haggard's experiment would show a left ear advantage in a

suitable non-speech context. Here, then, it is not the extraction of the acoustic cue which is important, but its phonetic relevance.

The existence of some stage which mediates between an acoustic representation of the input stimulus and the phonetic output has been suggested by Hiki *et al.* (1968) on the basis of experiments on a short-term contrast effect in vowel perception (Fry *et al.*, 1962). They suggest that there is some transform which maps acoustic space into a multi-dimensional phonetic space from which decisions are made about the appropriate phonetic category. The nature of this transform is determined both by the short-term effects that they investigated and by the longer term normalization effects demonstrated by Ladefoged and Broadbent (1957).

The arguments put forward here have concentrated on identifying the earliest stage at which differences between the ears become apparent. This is not necessarily the only stage, or that at which the greatest differences may be obtained. Work on temporal lobectomized patients has shown large differences between the two temporal lobes for verbal memory in excess of the short-term memory span (Milner, 1958), but there has been considerably less evidence that verbal perceptual deficits depend on which hemisphere is damaged. Luria (1966) presents some evidence that patients with damage to the left temporal lobe are impaired in their ability to repeat simple nonsense syllables. But this is the only evidence of its kind. The work on commissurectomized patients has given no evidence that there are any perceptual differences between the two hemispheres (Milner, Taylor and Sperry, 1968), although, of course, recall is largely restricted to only one hemisphere. Perceptual differences may in fact exist at the level of phonemic analysis, and these differences may not yet have been revealed because few tests have put strain specifically on the phonetic aspects of speech perception. That no effects, other than those reported by Luria, have yet appeared does suggest that the lateralization of speech perception is considerably less than that of speech production and verbal memory. This does not necessarily mean that these latter processes are influencing the results of the experiments reported here. It may well be that the dichotic listening technique is particularly sensitive to processes which occur early in the sequence of perception and memory, if only because stimuli are more likely to be differentiated according to ear of arrival immediately after input than at some later time. We must, however, acknowledge the possibility that memory processes may show differential ear effects, although there is yet little evidence that they do.

Appendix: *D*-Scores

In a free recall dichotic listening experiment, the simple percentage correct score is inadequate for two reasons. First, it takes no account of the relative number of times one ear is reported first and the other second, so that errors arising from serial order effects are confounded with those from other sources. Second, differences in percentage correct are not strictly comparable between subjects because of varying overall levels of performance; a given difference in detectability gives rise to a wide range of differences in percentage correct at different performance levels. The two *D* scores described here give estimates of the difference in recall between

the two ears on the first and second reported channels, respectively. These estimates take into account both the relative number of times each ear is reported first, and the absolute probability of being correct on each of these channels.

First and second channel here refer simply to the order of report rather than to any property of the input. The following letter combinations denote the number of trials on which each subject made the corresponding pattern of correct responses.

LR = left ear correct on first channel, right ear correct on second channel.
RL = right ear correct on first channel, left ear correct on second channel.
LZ = left ear correct on first channel, neither ear correct on second channel.
RZ = right ear correct on first channel, neither ear correct on second channel.
ZL = neither ear correct on first channel, left ear correct on second channel.
ZR = neither ear correct on first channel, right ear correct on second channel.
ZZ = neither ear correct on first channel, neither ear correct on second channel.

Then let:

$$p(\mathrm{L}_1) = (\mathrm{LR} + \mathrm{LZ})/(\mathrm{LR} + \mathrm{LZ} + \mathrm{ZR})$$
$$p(\mathrm{R}_1) = (\mathrm{RL} + \mathrm{RZ})/(\mathrm{RL} + \mathrm{RZ} + \mathrm{ZL})$$
$$p(\mathrm{L}_2) = (\mathrm{RL} + \mathrm{ZL})/(\mathrm{RL} + \mathrm{RZ} + \mathrm{ZL})$$
$$p(\mathrm{R}_2) = (\mathrm{LR} + \mathrm{ZR})/(\mathrm{LR} + \mathrm{LZ} + \mathrm{ZR})$$

Denoting a normal transformation with a prime we now define

$$D_1 = p'(\mathrm{R}_1) - p'(\mathrm{L}_1)$$
$$D_2 = p'(\mathrm{R}_2) - p'(\mathrm{L}_2)$$

This scoring method ignores trials on which neither ear was correct (ZZ), and assumes that making a normal transformation is an adequate compensation for variations in overall performance level (Green and Birdsall, 1964).

The first experiment reported here was performed at the Psychological Laboratory, Cambridge, under the supervision of Dr D. E. Broadbent, while I held a Medical Research Council Studentship. I was also supported by the Commonwealth Fund and by a grant from the National Institute of Child Health and Development to Haskins Laboratories.

This paper would not have been possible without the help of Drs M. P. Haggard, A. M. Liberman, I. G. Mattingly, D. P. Shankweiler, M. Studdert-Kennedy and T. G. Halwes. Professor O. L. Zangwill and Dr F. S. Cooper generously provided research facilities.

References

BAKKER, D. J. (1968). Ear-asymmetry with monaural stimulation. *Psychonomic Science*, **12**, 62.

BAKKER, D. J. (1970). Ear-asymmetry with monaural stimulation: relations to lateral dominance and lateral awareness. *Neuropsychologia*, **8**, 103–17.

BROADBENT, D. E. (1958). *Perception and Communication*. London: Pergamon.

BROADBENT, D. E. and GREGORY, M. (1964). Accuracy of recognition for speech presented to the right and left ears. *Quarterly Journal of Experimental Psychology*, **16**, 359–60.

BRYDEN, M. P. (1963). Ear preference in auditory perception. *Journal of Experimental Psychology*, **65**, 103–5.

BRYDEN, M. P. (1967). An evaluation of some models of laterality effects in dichotic listening. *Acta oto-laryngologica*, **63**, 595–604.

CHIBA, T. and KAJIYAMA, M. (1941). *The Vowel—Its Nature and Structure*. Tokyo.

CORSI, P. M. (1967). The effects of contralateral noise upon the perception and immediate recall of monaurally presented verbal material. Unpublished M.A. thesis. McGill University, Montreal.

DARWIN, C. J. (1969). Auditory perception and cerebral dominance. Unpublished Ph.D. thesis, University of Cambridge.

FANT, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *STL-QPSR*, **4**, 22–30.

FRY, D. B., ABRAMSON, A. S., EIMAS, P. D. and LIBERMAN, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, **4**, 171–89.

FUJISAKI, J. and KAWASHIMA, T. (1968). The roles of pitch and higher formants in the perception of vowels. *Institute of Electrical and Electronic Engineers Transactions*, AU-16, 73–7.

GREEN, D. M. and BIRDSALL, T. M. (1964). The effect of vocabulary size on association score. In SWETS, J. A. (Ed.) *Signal Detection and Recognition by Human Observers*, pp. 609–19. New York: Wiley.

HAGGARD, M. P. (1971a). Encoding and the REA for speech signals. *Quarterly Journal of Experimental Psychology*, **23**, 34–45.

HAGGARD, M. P. (1971b). Stimulus and task factors as determinants of ear advantages. *Quarterly Journal of Experimental Psychology*, **23** (in press).

HAGGARD, M. P., AMBLER, S. and CALLOW, M. (1969). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, **47**, 613–7.

HALL, J. L. and GOLDSTEIN, M. H. (1968). Representation of binaural stimuli by single units in primary auditory cortex of unanaesthetized cats. *Journal of the Acoustical Society of America*, **43**, 456–61.

HALWES, T. G. (1969). Effects of dichotic fusion on the perception of speech. *Supplement to Status Report on Speech Research*, Sept. 1969, Haskins Labs.

HARRIS, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, **1**, 1–17.

HEINZ, J. M. and STEVENS, K. N. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, **33**, 589–96.

HIKI, S., SATO, H. and OIZUMI, J. (1968). Dynamic model of vowel perception. Paper presented to 6th International Congress of Acoustics, Tokyo, Japan, August 1968.

HYDE, S. R. (1968). *Automatic Speech Recognition*. Post Office Research Dept. Report No. 45.

INGLIS, J. (1962). Dichotic stimulation, temporal lobe damage, and the perception and storage of auditory stimuli—a note on Kimura's findings. *Canadian Journal of Psychology*, **16**, 11–6.

KIMURA, D. (1961*a*). Some effects of temporal lobe damage on auditory perception. *Canadian Journal of Psychology*, **15**, 156–65.

KIMURA, D. (1961*b*). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, **15**, 166–71.

KIMURA, D. (1964). Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology*, **14**, 355–8.

KIMURA, D. (1967). Functional asymmetries of the brain in dichotic listening, *Cortex*, **3**, 163–78.

KIRSTEIN, E. and SHANKWEILER, D. P. (1969). Selective listening for dichotically presented consonants and vowels. Paper read before 40th Annual Meeting of Eastern Psychological Association, Philadelphia.

LADEFOGED, P. and BROADBENT, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, **29**, 98–104.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P. and STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431–61.

LURIA, A. R. (1966). *Human Brain and Psychological Processes*. Trans. B. Haigh. New York and London: Harper and Row.

MATTINGLY, I. G. (1966). Speaker variation and vocal-tract size. *Journal of the Acoustical Society of America*, **39**, 1219 (A).

MATTINGLY, I. G. (1968). Experimental methods for speech synthesis by rule. *Institute of Electrical and Electronic Engineers Transactions*, AU-**16**, 198–202.

MILLER, G. A. and NICELY, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, **27**, 338–52.

MILNER, B. (1958). Psychological deficits produced by temporal lobe excision. *Publications of the Association for Research in Nervous Mental Diseases*, **36**, 244–57.

MILNER, B. (1962). Laterality effects in audition. In MOUNTCASTLE, V. B. (Ed.), *Interhemispheric Relations and Cerebral Dominance*. Baltimore: Johns Hopkins.

MILNER, B., TAYLOR, L. B. and SPERRY, R. W. (1968). Lateralized suppression of dichotically-presented digits after commissural section in man. *Science, New York*, **161**, 184–5.

OXBURY, S., OXBURY, J. and GARDINER, J. (1967). Laterality effects in dichotic listening. *Nature, London*, **214**, 742–3.

ROSENZWEIG, M. R. (1951). Representation of the two ears at the auditory cortex. *American Journal of Physiology*, **167**, 147–58.

SHANKWEILER, D. P. (in press). An analysis of laterality effects in speech perception. In KJELDERGAARD, P. (Ed.), *The Perception of Language*. Pittsburgh, Pa.: University of Pittsburgh Press.

SHANKWEILER, D. P. and STUDDERT-KENNEDY, M. (1967*a*). Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, **19**, 59–63.

SHANKWEILER, D. P. and STUDDERT-KENNEDY, M. (1967*b*). An analysis of perceptual confusions in identification of dichotically presented CVC syllables. *Journal of the Acoustical Society of America*, **41**, 1581 (A).

SIEGEL, S. (1956). *Non-parameteric Statistics*. New York: McGraw Hill.

SPARKS, R. W. and GESCHWIND, N. (1968). Dichotic listening in Man after section of neocortical commissures. *Cortex*, **4**, 3–16.

STUDDERT-KENNEDY, M. and SHANKWEILER, D. P. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, **48**, 579–94.

TREISMAN, A. M. and GEFFEN, G. (1968). Selective attention and cerebral dominance in perceiving and responding to speech messages. *Quarterly Journal of Experimental Psychology*, **20**, 139–50.

TUNTURI, A. R. (1946). A study of the pathway from the medial geniculate body to the acoustic cortex in the dog. *American Journal of Physiology*, **147**, 311–9.

WICKELGREN, W. A. (1966). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, **39**, 388–98.

Received 28 September 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 63–71

NATURAL AND CONTRIVED EXPERIENCE IN A REASONING PROBLEM

P. C. WASON AND DIANA SHAPIRO

Psycholinguistics Research Unit, Department of Phonetics, and Department of Psychology, University College London

This study is concerned with the effects of prior experience on a deceptive reasoning problem. In the first experiment the subjects (students) were presented with the problem after they had experienced its logical structure. This experience was, on the whole, ineffective in allowing subsequent insight to be gained into the problem. In the second experiment the problem was presented in "thematic" form to one group, and in abstract form to the other group. Ten out of 16 subjects solved it in the thematic group, as opposed to 2 out of 16 in the abstract group. Three hypotheses are proposed to account for this result.

Introduction

This study is about the effects of two kinds of experience on a deceptive reasoning problem. In the first experiment the experience is introduced as part of the procedure, and in the second it is inherent in the material used.

Previous experiments (Wason, 1968, 1969*a*) have established that it is very difficult to decide what information is required to test the truth of an abstract conditional sentence. For example, given the sentence: *Every card which has a D on one side has a 3 on the other side* (and knowledge that each card has a letter on one side and a number on the other side), together with four cards showing respectively D, K, 3, 7, hardly any individuals make the correct choice of cards to turn over (D and 7) in order to determine the truth of the sentence. This problem is called the "selection task" and the conditional sentence is called "the rule".

The rule has the logical form, "if p then q", where p refers to the stimulus mentioned in the antecedent (D); $\bar{p}$, i.e. not p, refers to the stimulus which negates it (K); q refers to the stimulus mentioned in the consequent (3); and $\bar{q}$, i.e. not q, refers to the stimulus which negates it (7). In order to solve the problem it is necessary and sufficient to choose p and $\bar{q}$, since if these stimuli were to occur on the same card the rule would be false but otherwise true.

The combined results of four experiments (see Table I) show that the subjects (students) are dominated by *verification* rather than *falsification*. On the whole, they failed to select $\bar{q}$, which could have falsified the rule, and they did select q, which could not have falsified it although this latter error is much less prevalent.

Experiment I

The previous experiments have been concerned with the stability of the errors and their resistance to correction by "remedial procedures". After the subjects had performed the selection task they had to evaluate the cards independently,

i.e. turn them over and say whether the rule was true or false in relation to each. The present experiment is concerned with the prevention of error. The subjects are made familiar with the other side of the cards before the selection task is performed.

This prior experience is introduced by two methods. The "construction" method requires the subject to imagine, or project, a value on the other side of a card which would make the rule true, or make it false, in relation to it. In effect, positive and negative instances of the rule are constructed. The "evaluation" method simply requires the subject to turn over the card and say whether the rule is true, or false, in relation to it. The construction method clearly involves an imaginative act, and hence a greater degree of involvement than the evaluation method. It was accordingly predicted that it would be associated with superior performance in the subsequent selection task.

TABLE I

Frequency of the selection of cards in four previous experiments (n = 128)

p and q	59
p	42
p, q and $\bar{q}$	9
p and $\bar{q}$	5
others	13

Design

Two independent groups were used: the construction group and the evaluation group. Both carried out their respective tasks on 24 cards in relation to a given conditional rule. They then performed the *initial selection task* with four more cards in relation to the *same* rule. A new conditional rule was then presented together with a further four cards. This *transfer selection task* was designed to assess the extent to which specific knowledge, gained in the prior experience, would be generalized.

Subjects

Twenty-four undergraduates (paid volunteers) of University College London were allocated alternately to the groups and tested individually. They had no previous experience with tasks of this type.

Procedure

Before presenting the rule all the subjects were first handed 28 cards, and instructed to inspect them to ensure that each had a letter of the alphabet on one side and a number on the other side.

They were then presented with the following rule: *Every card which has a vowel on one side has an even number on the other side.* Twenty-four of the 28 cards were then presented, one at a time, the remaining four being reserved for the transfer selection task. In the construction group they were instructed to name a value on the other side of each card which would make the rule true (or make it false). They were, however, told that it would be in order to say that no value on the other side would make the rule either true or false.

In the evaluation group they turned over each card and said that it made the rule true (or false). Similarly, they were told that it would be in order to say that a card was irrelevant to the truth or falsity of the rule.

The eight possible ways of permuting the logical values were each represented three times in the series of 24. They were presented successively in the following pairs, where the value given first refers to the symbol uppermost: $(pq, p\bar{q})$ $(\bar{p}\bar{q}, \bar{p}q)$ $(qp\ q\bar{p})$ $(\bar{q}\bar{p}, \bar{q}p)$. All the subjects received the cards in the same order, and within a pair the order of presenting the two cards was constant, but the pairs themselves were randomized in a different order within each of the three blocks of eight cards. In the construction group, where only the uppermost symbol was presented, the instruction for the first card within a pair was to name a symbol to make a verifying instance, and for the second card to name a symbol to make a falsifying instance.

In both groups the subjects were told they were wrong if they failed to evaluate (construct) $p\bar{q}$ and $\bar{q}p$ as falsifying, and if they did evaluate (construct) $\bar{p}q$ and $q\bar{p}$ as falsifying. This was to ensure that they did appreciate the falsifying instances of a conditional rule, but did not confuse them with the falsifying instances of an equivalence rule. The $\bar{p}q$ and $q\bar{p}$ instances do falsify an equivalence rule in the form: "if, and only if p then q".

For the initial selection task four cards (E, Z, 6, 7), taken from the 24 used in the prior experience, were placed on the table in a random order. The subjects were instructed that the rule now applied to these four cards taken as a whole, i.e. no longer independently. They were told "to select those cards, and only those cards, that would need to be turned over in order to discover whether the rule was true or false". No comments were made about these selections, and the subjects were not allowed to turn over any of the cards.

For the transfer selection task the following rule was presented: *Every card which has a D on one side has a 3 on the other side*, together with the four cards (D, K, 3, 5) which had not occurred in the series of 24, but had been included in the 28 originally inspected. The instructions were similar to those given for the initial selection task.

Results

Table II shows the frequency of correct and incorrect solutions, the first number in each cell referring to the initial selection task and the second to the transfer selection task.

TABLE II

Frequency of correct and incorrect solutions

	Correct	Incorrect	*N*
Construction	5 (6)	7 (6)	12
Evaluation	2 (2)	10 (10)	12
Totals	7 (8)	17 (16)	24

As predicted, there is a trend in favour of the construction group, but it falls short of statistical significance. The performance overall is unimpressive, particularly in the evaluation group. It will also be noted that the difference between the two selection tasks is negligible: knowledge is generalized to the extent that it has been gained. The two types of error, i.e. the selection of q and the omission of $\bar{q}$ are examined separately in Table III and IV.

Table III shows that both groups do better in omitting q than in getting the solution correct. But the frequency of this particular error also increases the

difference between the groups in the predicted direction. On the transfer selection task it is significant ($P = 0{\cdot}05$, one-tailed, Fisher-Yates exact test).

It may be inferred from Table IV that the proportion of subjects in both groups who select $\bar{q}$ is greater than ever obtained initially in previous experiments. But it is also evident that none of the frequencies differ from chance expectancy. However, it may be inferred that the trend, showing the construction group superior on the correct solution, is entirely due to a greater tendency to omit q rather than one to select $\bar{q}$. The reasons for this result must be sought in the two different methods of introducing prior experience.

TABLE III

Frequency of selecting q

	q selected	q omitted	N
Construction	2 (3)	10 (9)	12
Evaluation	7 (8)	5 (4)	12
Totals	9 (11)	15 (13)	24

TABLE IV

Frequency of omitting $\bar{q}$

	$\bar{q}$ omitted	$\bar{q}$ selected	N
Construction	7 (5)	5 (7)	12
Evaluation	6 (6)	6 (6)	12
Totals	13 (11)	11 (13)	24

In the evaluation group the responses made in the prior experience corresponded to the "defective truth table" which is followed when a conditional sentence is evaluated (Johnson-Laird and Tagart, 1969). According to this truth table pq is classified as true, $p\bar{q}$ as false, and both $\bar{p}q$ and $\bar{p}\bar{q}$ as irrelevant. Only 1·4% of the evaluations deviated from this classification. In contrast, the truth table for the conditional in the propositional calculus counts all contingencies as true except $p\bar{q}$.

A very different picture emerges in the construction group. Table V shows the frequencies of instances constructed over the first block of eight trials, and (in parentheses) over the third block of trials.

Inspection of Table V suggests some tendency to reason by equivalence initially, i.e. to construct pq, qp, $\bar{p}\bar{q}$, $\bar{q}\bar{p}$, as verifying instances, and $p\bar{q}$, $\bar{q}p$, $\bar{p}q$, $q\bar{p}$, as falsifying instances. A third of the subjects were consistent in interpreting all the contingencies in this way over the first block of trials, and there was a very slight tendency for these subjects to perform better on the selection tasks. However, as a function of the feedback that $\bar{p}q$ and $q\bar{p}$ do not falsify, it will be observed that in the third block of trials there was a strong tendency to deny these contingencies falsifying status.

TABLE V

Frequencies of instances constructed on the first and third blocks of trials (n = 12)

Value presented		Value constructed p	$\bar{p}$	q	$\bar{q}$	none
p	T?			12 (12)		0 (0)
	F?				12 (12)	0 (0)
$\bar{p}$	T?			1 (3)	7 (4)	4 (5)
	F?			<u>6 (0)</u>	1 (0)	5 (12)
q	T?	11 (10)				1 (2)
	F?		<u>7 (1)</u>			5 (11)
$\bar{q}$	T?		8 (7)			4 (5)
	F?	<u>9 (10)</u>	1 (0)			2 (2)

T = true, F = false. The numbers in parentheses refer to the third block of trials. The contingencies affected by feed-back are underlined.

This departure from the defective truth table may have been because the procedure was taken as a challenge to construct an instance. The subjects may have adopted a weaker standard of truth for the verifying instances, i.e. mere consistency with the rule. But when they proceeded to construct $\bar{p}q$ and $q\bar{p}$ as falsifying instances, they would have been corrected. The irrevelance of q would then have been learned directly, and this is reflected in the performance of the construction group in the selection tasks (see Table III).

It seems much more surprising that in both groups only about half the subjects selected $\bar{q}$ (see Table IV). An information-processing model, which has been devised to explain performance in these tasks (Johnson-Laird and Wason, 1970*a*), elucidates this result. The model postulates that the recognition that $\bar{q}$ could falsify the rule is not, in itself sufficient for its selection. Its relevance is assumed to depend on the arousal of a conflict between pq (as verifying) and $\bar{q}p$ (as falsifying). The conditions for this conflict occur if p had been selected and $\bar{q}$ omitted. But in the present experiment these conditions would be unlikely to have occurred because successive instances were constructed, or evaluated, independently of each other before the selection task is performed.

It may be concluded that the putative experience of logical structure, introduced procedurally, is relatively ineffective in enabling insight to be gained into the problem. It is reasonable to enquire whether "natural" experience, inherent in the subjects' everyday knowledge, may be more successful in inducing insight. It was predicted that when the material is realistic ("thematic"), as opposed to abstract, the selection task will be significantly easier.

Experiment II

Design

Two independent groups were used: the "thematic group" and the "abstract group" which differed solely in the terms in which the problem was presented.

Subjects

Thirty-two first year psychology undergraduates of University College London were allocated alternately to the groups and tested individually. They had no previous experience with tasks of this type.

Procedure

The thematic material represented a journey made on 4 different days of the week. Before presenting the rule about these journeys the subjects were given 16 cards which they inspected to ensure that each had the name of a town on one side and a mode of transport on the other side.

They were then presented with the four selection task cards, taken from the 16 originally presented, and arranged in random order on the table. They were instructed that they would now only be concerned with these cards. On two of them a different destination was written, i.e. "Manchester" and "Leeds", and on the other two a different mode of transport, i.e. "Car" and "Train". In addition each had a different day of the week in smaller type at the top.

The rule was then presented as a claim made by the experimenter about four journeys she had made on the four different days indicated on the cards. One variant of this rule was: *Every time I go to Manchester I travel by car.* Three other variants, derived from permuting the items on the cards, were also used. The presentation of all four was systematically rotated between the subjects to control for any possible preconceptions about the relation between destinations and modes of transport.

It was explained to the subjects that for each journey the destination appeared on one side of the card and the transport used on the other side. They were then instructed to say which cards they would need to turn over to decide whether the experimenter's claim was true or false. They were encouraged to take their time before answering.

A similar procedure was followed in the abstract group. Sixteen cards with a letter of the alphabet on one side and a number on the other side were first inspected. Four of these, D, K, 3, 7, were used for the selection task. The rule: *Every card which has a D on one side has a 3 on the other side*, was then presented as a claim made by the experimenter about the arrangement of letters and numbers on the cards. The subjects were instructed that this rule applied only to the four cards, and that they were to say which they would need to turn over to decide whether the claim was true or false.

Results

Table VI shows the frequency of correct and incorrect solutions.

TABLE VI

Frequency of correct and incorrect solutions

	Correct	Incorrect	N
Thematic	10	6	16
Abstract	2	14	16
Totals	12	20	32

The prediction that the thematic group would perform better than the abstract group is clearly confirmed by the distribution of the frequencies in Table VI ($P = 0{\cdot}004$, one-tailed, Fisher-Yates exact test). It is evident that representing

the problem in the form of a realistic situation had a dramatic effect on the subjects' ability to gain insight into it. There may, however, be several reasons for this result.

Discussion

The results of the two experiments show the relative failure of procedurally introduced experience and the relative success of realistic material in allowing insight to be gained into the problem.

It could, of course, be argued that if the experience, introduced in Experiment I, had been more intensive, or if only the falsifying contingencies had been used, then performance would have been improved. But the purpose of the experience was only to acquaint the subjects with the logical structure of the problem, and not to train them to make particular responses. Previous results (Johnson-Laird and Wason, 1970*b*) have shown that various factors, such as cognitive load, may affect the appreciation of the task, and over-learning of the contingencies might be one more variable affecting performance. The point is that understanding the contingencies did not allow this knowledge to be used with maximum efficiency in the selection tasks. This result may seem incredible to anyone unacquainted with the difficulty of the problem. The reasons for it will not be discussed until the effects of thematic material on the task have been considered because these help to explain it.

Three hypotheses about different aspects of the thematic material used in Experiment II could account for its beneficial effects. First, the terms used in the thematic material, the towns and modes of transport, are concrete as opposed to the abstract terms which consisted of letters and numbers. It is well known that concrete material is better remembered than abstract material, and that in syllogistic reasoning familiar terms inhibit fallacious inferences (Wilkins, 1929). Thus in Experiment II the concrete terms may have been symbolically manipulated more readily and more appropriately than the abstract terms. This hypothesis might be tested by using concrete terms with an arbitrary connection, e.g. "Every card which has *iron* on one side has *apple* on the other side", where metals and fruits are known to occur on either side of the cards.

Second, it may be the concrete relation between the terms, rather than the terms themselves, which is beneficial. In the thematic material the relation which connects the terms is "travelling", as opposed to "the other side of the card" which connects the abstract material. This hypothesis could be tested by using abstract terms with a concrete relation between them, e.g. "Every time I go to K I travel by 3", where letters and numbers are known to stand for towns and transport respectively.

Third, the thematic material, unlike the abstract material, forms a coherent, unified whole: a claim about journeys supposed to have been made on four different days. Hence the subjects may have been more inclined to distribute their attention equally on its components, i.e. the four cards. They would thus be liberated from fixations on those cards which correspond to items mentioned in the rule. Cyril Burt (personal communication) has even suggested that thematic

material enables the subjects to concentrate on the situation depicted, unfettered by the presence of the cards. This does not, in itself, explain why thematic material is helpful. But if it is assumed that knowledge about such material is represented in the brain in schemata, which may be activated by appropriate cues, then the solution to the problem may be simply "read off" by reference to this stored information.

The abstract material has no unifying link: each card is distinct and separate rather than being parts of a whole. The subjects are instructed that the rule refers only to the four cards, but in spite of this they may have construed it merely as a formula. They may, in fact, have regarded the cards as items in a sample from a larger universe, and reasoned about them inductively rather than deductively. In doing this they may have implicitly followed the Bayesian rule which assumes that the probability of a generalization is increased by repetition of confirming instances. Hence they might not have been disposed to consider the potential relevance of $\bar{q}$. There was some introspective support for probabilistic reasoning of this kind. It would follow, of course, that the experience of the problem's logical structure, introduced in Experiment I, would not have disabused the subjects of this particular misconception.

In fact, the difficulty of the abstract selection task may be due, not to the failure to recognize the correct solution, but to the failure to generate alternatives in order to derive the correct solution. In other words, abstract material may inhibit the realization of the necessity of combinatorial analysis rather than hindering the performance of such an analysis. The meaninglessness of the rule may tempt the subjects to interpret it, not as a rule, but as a sentence to be matched against instances. With thematic material it is gratuitous to talk about combinatorial analysis: the activation of stored knowledge spontaneously generates "real" alternatives. This hypothesis might be tested by comparing thematic and abstract material, but presenting all the possible solutions in a list from which one has to be selected, thus obviating the need for a combinatorial analysis. It would then be predicted only that the correct solution would be located more quickly with thematic material than with abstract material without a difference in its relative frequency.

Finally, the present results support the suggestion (Wason, 1969*b*) that it is not so much the logical structure which makes the abstract problem difficult, as the structure which the subjects impose upon the problem. Its difficulty does not lie in the fact that inferences of the kind demanded "hardly ever occur in real life"—a criticism sometimes voiced of the early experiments. On the contrary, when the task is made realistic it becomes appreciably easier. What makes the abstract task difficult is the arbitrariness of material which seems to defy the reasoning process. A more precise definition of the impediments involved must await further investigation.

The experiments in this paper form part of research to be reported in a thesis to be submitted for the degree of Ph.D. of London University by the second author, under the supervision of the first author. We are most indebted to our colleague, Dr P. N. Johnson-Laird, for invaluable critical comments and suggestions, and also to the Medical Research Council for a grant for scientific assistance.

References

JOHNSON-LAIRD, P. N. and TAGART, J. (1969). How implication is understood. *American Journal of Psychology*, **82**, 367–73.

JOHNSON-LAIRD, P. N. and WASON, P. C. (1970*a*). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology* **1**, 134–48.

JOHNSON-LAIRD, P. N. and WASON, P. C. (1970*b*). Insight into a logical relation. *Quarterly Journal of Experimental Psychology*, **22**, 49–61.

WASON, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 273–81.

WASON, P. C. (1969*a*). Regression in reasoning? *British Journal of Psychology*, **60**, 471–80.

WASON, P. C. (1969*b*). Structural simplicity and psychological complexity. *Bulletin of the British Psychological Society*, **22**, 281–84.

WILKINS, M. C. (1929). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, No. 102.

Received 23 *April* 1970

necessarily predict the cue tone, but that implicitly following instructions was of more importance than attempting to achieve a high score. Cue sequences (constrained) and row instructions were so combined that in each set of 10 exposures the subject looked 4 times in the correct position and once in each of the 6 possible wrong positions.

Results

Eye movement data

Measurements of eye position are notoriously difficult to make and, as the visual angle of the display entailed operating the equipment near its limit, no great accuracy was expected in the recordings. Of the 360 exposure instances, eye position was deemed measurable in 236 cases which were used for analysis. For these cases eye position at the onset of each exposure was estimated as falling into one of five locations, corresponding to the high, middle and low letter rows and the two intermediate positions. In several cases subjects exhibited the behaviour described by Bryden (1961), flicking the eyes towards the correct row on discriminating the cue tone, or engaged in blinks following the stimulus exposures; however, only the point fixated at the onset of the flash was recorded.

The distribution of estimated eye positions, shown in Table I, diverges markedly

TABLE I

Eye fixations at different row locations

H	MH	M	ML	L	Total
30	81	96	14	15	236

from even usage of the five positions. The divergence is of course highly significant ($\chi^2 = 126$, $P < 0{\cdot}001$) and takes the form of an accentuation of the middle at the expense of the other locations, with a bias towards the high row which accords well with the guessing data of the previous experiment. It is probable that preference for the middle positions is largely a side effect of the measurement technique, since the recurrent demands to return the gaze to the middle line for calibration purposes will strongly induce a set to fixate that line on exposure. If this is so the tendency to anticipate, which it was hoped to measure, will be much reduced.

In fact, a change of this kind does appear to have occurred. The differences in accuracy between the random, constrained and non-random conditions are found significant and in the predicted order ($P < 0{\cdot}05$) by Page's (1963) L test; but the slope of the increase with degree of statistical constraint is reduced in comparison with the previous experiment, presumably because the tendency to fixate middle rows prevented subjects from taking advantage of non-randomness. Both the present and previous sets of scores are shown in Table II; interference by the eye movement apparatus is probably at least partly responsible for the lower overall level of accuracy.

Plotting for each condition the percentage of times that eye fixations are in the correct or variously incorrect positions yields similar results. The proportion of

TABLE II

Mean number of letters correct with differing cue predictability

	Random	Constrained	Non-random
Previous experiment	1·83	2·03	2·42
Present data	1·37	1·50	1·78

completely correct fixations is somewhat higher for the non-random than for the combined random and constrained conditions ($z = 2{\cdot}2$, $P < 0{\cdot}02$); but the largest proportion of fixations for all conditions is at one row distant from the correct location, according with the preference for fixation of the middle row.

The most important data are obtained by plotting mean letter scores, against the degree of separation between the row location required for partial report and the point of eye fixation. This curve is shown in Figure 1, superimposed upon the

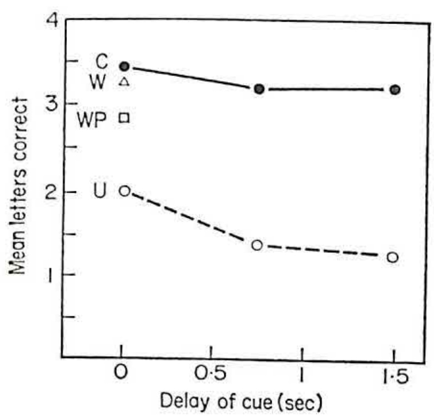

FIGURE 1. Dependence of amount recalled upon distance between attended row and cued row.

results from the previous experiment and from the second stage of the present one. Despite the relative unreliability of the eye movement data, the trend is in excellent correspondence with the other two sets of data, showing a progressive decline in the number of letters reported as the point of fixation is removed from the cued row.

Effects of advance instruction

It is clearly open to subjects to attempt an increase in score by making use of intermediate fixation points and by developing mixed or "hedging" strategies, which will normally be more successful with closely spaced letter rows than with wider spacing. Thus, provided the subject has no subsequent access to a persisting image, the difference between widths of spacing should be much reduced when the subject is directed to look at single rows. The mean letter scores for day 3, shown in Table III, show no significant difference between spacings (Wilcoxon; $T = 15$, $N = 9 + 3$ ties) when mixed strategies are precluded by instructions, but a significant difference ($P < 0{\cdot}05$) when subjects are run under normal partial report conditions.

When subjects are instructed where to look their gaze may be directed at the called row, one or two spacing units away on the single-spaced cards, or two or four spacing units away in the case of the double-spaced cards. Zero separation is given by looking high when "high" is to be reported, or H(H), and by M(M) and L(L) on both sets of cards; a separation of one unit is represented by H(M), M(H), L(M) and M(L) in single-spacing; two units are given by H(L) and L(H) (single-spacing) and by H(M), M(H), L(M) and M(L) (double-spacing); while H(L) and L(H) (double-spacing) are separated by four units. The mean letter scores for each degree of separation are shown in Figure 1. Data from both sets of cards were combined for testing by Page's L, which shows the scores to differ significantly ($P < 0{\cdot}001$) between separation levels in the order $1 > 2 > 3 > 4$. The agreement seen in Figure 1 between the three sets of data is striking. Attempting to determine where subjects are looking by recording their guesses, by measuring eye position and by giving advance directions, in each case suggests a sharp decline in the amount seen as distance from the point of fixation increases.

TABLE III

Mean letters correct as a function of row spacing and instructions

Spacing	Normal procedure	Instructed gaze
Narrow	1·88	1·65
Wide	1·48	1·59

The steepness of the slope is probably underestimated, as similar trends arise from the typical errors of all three methods. The effects of looking at a row not announced as a guess, of hesitation in scoring an eye position at the scale extremes, and of not looking where directed, will all tend to increase the apparent scores at the wider separations while reducing those closer to the target row. Similarly, all three methods identify only horizontal rows, so that scoring on the wider separations will be inflated, and scoring on the narrower separations depressed. on each occasion that a subject attends to a vertical segment of the array. Nevertheless, the present data appear clearly to differ from the flat curve which would presumably be required, at least as far as two separation units, in order to justify the assumption of sampling from a visual image.

Experiment II

It appears that subject typically has access to very few letters, but he may distribute these letters within the array in different ways in order to maximize his partial report score. Thus, depending upon the sequence of cues and the dispersion of the array, he may choose to select complete rows, a block of letters in adjoining rows, or vertical columns of letters; the latter techniques are possible bases for the "equal-attention" strategy. These considerations appear sufficient to account for reporting behaviour when the cue follows immediately upon presentation of the array. However, if it is assumed that a subject making a partial

report does not refer back to a visual image, the effects due to delaying the cue tone remain to be explained.

It was shown (Sperling, 1960; Experiment 3) that partial report scores decline systematically as the onset of the cue tone is delayed. On the hypothesis that more letters are somehow available to the subject than can be accommodated by the memory span, and that the extra capacity is supplied by a temporary image, it is natural to ascribe the decline in score to degradation of the image. However, on the assumption that the subject has access to only a few letters, the amount to be retained will normally fall within the memory span, so that a different explanation of the delay effects is required. A promising explanation may be derived from a comparison of the two reporting procedures.

The ostensible difference between procedures is that the whole report attempts to enumerate, while the partial report samples, what is available. However, the partial report also differs from the whole report in that an extra information load due to discriminating the cue tone is imposed upon the subject during the retention period, and in that the partially reporting subject is uncertain what to select for observation and what to rehearse. These two factors are usually confounded, and probably have interacting effects. When separated out, it seems probable that the extra discrimination load is of minor importance with the two or three cue alternatives currently used, despite the fact that the discrimination problem rapidly becomes exacerbated if more complex partial reporting procedures are tried. Thus, combining whole report with an information load in the form of a two-choice cue for order of report, in Sperling's (1960; Experiment 7) data, gives rise to little decline with delay.

The most important factor therefore appears to be uncertainty. Uncertainty about what is to be reported will directly affect scores with instantaneous cueing, but will also tend to generate inefficient rehearsal strategies. The subject will vacillate, for instance, between stressing the seen items of an expected middle row, repeating a vertical sequence for safety, or withholding rehearsal while attempting to reconstruct partially seen letters. Inefficiency of rehearsal will lead to a progressive decline in the number of letters available for delayed report, the decline becoming greater as uncertainty increases. Thus, while selection uncertainty should affect the intercept, rehearsal uncertainty should affect the slope of the decay curve.

The experiment to be described explores the effects of uncertainty in two ways. With cue load present, as always in partial report conditions, cue delay is varied for two levels of uncertainty. With cue load absent as in whole report conditions varying delay of report is impracticable, but two levels of uncertainty are manipulated in a way which makes possible some further examination of rehearsal methods.

Method

Subjects and apparatus

As skilled observers are desirable, 11 of the 12 previously used subjects were recalled; the data were supplemented by testing one additional subject who had earlier been used in informal trials. The apparatus was as described for the previous experiment. All testing was carried out with the single-spaced cards.

Procedure

Each subject made written, partial reports on sets of 10 stimulus cards at cue delays of 0, 0·75 and 1·5 sec, at two levels of uncertainty:

Uncertain. The uncertainty was generated by the properties of the constrained cue sequences. These were used in preference to the random sequences, which give rise to scores so low as to impede the measurement of any decline, and may encourage stereotyped rehearsal strategies since no method is better than chance.

Certain. The non-random sequences give rise to relatively uncertain but variable behaviour, when subjects do not know that they are non-random. The method of telling subjects beforehand which cue to expect was therefore adopted; in the earlier guessing experiment this technique was shown to give results quantitatively consistent with those under the other cue sequence constraints.

In addition to the 6 partial report combinations, each subject made reports under two conditions in which the cue tone was absent:

Whole. The instruction was to attempt to write down as many letters as possible in the correct positions, regardless of the row from which they were drawn.

Whole/partial. Subjects were again instructed to write down as many letters as possible but were warned that only one of the rows, to be designated after each recall attempt, would be scored. The object of this instruction was to inject the uncertainty of the normal, partial report procedure into the whole report situation, without imposing a cue load during rehearsal. Random lists were used for the sequences of row information.

The two whole conditions were run alternately at the beginning and end of the testing sessions, bracketing a counterbalanced presentation of the partial report conditions.

Results

Delay curves

The mean scores for the two partial report conditions are represented in Figure 2. In the case where anticipation is based on accurate foreknowledge, the certain

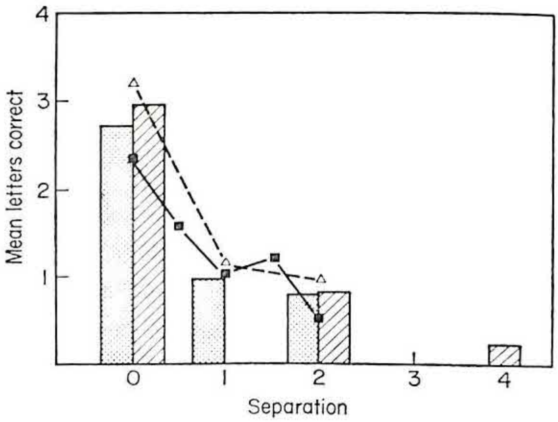

FIGURE 2. Amount recalled as a function of delay of cue and degree of uncertainty. Exp. I: eye movement records (■-■). Exp. II: Instructed gaze—single spacing (⬚); double spacing (▨). Previous exp. guesses (△ - - △).

condition shows no appreciable (or significant) drop with cue delay. In the uncertain condition, where the subject does not know what or how to rehearse, the effect of delay is to produce a deterioration in the recall scores. Page's L confirms ($P < 0·01$) that the letter scores at 0 sec delay are higher than those at 0·75 sec, which in turn exceed the scores at 1·5 sec. The slope appears less than in the corresponding graphs from Sperling (1960), but the apparent discrepancy is

an artefact due to the technique of multiplying the latter figures by 3 in order to estimate the number of letters "available."

Whole reports

The mean score for the standard whole report is 3·25 letters. Reference to Figure 2 shows that this appears to be slightly lower than the zero delay score when subjects are forewarned in the certain condition. The implication from a difference of this kind would be that subjects under self-instruction are less decisive in their selection than when unambiguously instructed by the experimenter, with a corresponding effect upon the efficiency of recall. The difference is not significant, but is in the same direction as a comparable effect in the previous work (Holding, 1970) and may therefore deserve attention.

The data from the whole/partial procedure may be used to derive two different scores. Treated as a partial report score, in the manner used when giving feedback to subjects, the mean letter score was 1·09. The main interest, however, lies in scoring the records as whole reports. In this case, the mean number of letters correctly reported is 2·84; this represents a significant drop (Wilcoxon; $P < 0{\cdot}05$) from the number correct under the standard whole procedure. Thus, although in both conditions subjects report as many letters as possible, the effect of uncertainty is to reduce the number correct.

Recall patterns

The notion that the uncertainty of the whole/partial procedure generates more disorganized rehearsal strategies may be examined in greater detail by considering the spatial arrangement of recall attempts.

In the whole condition, subjects typically attempted 4·08 correct and incorrect letters, normally in single rows with occasional extra letters. Of the 120 recall trials, only 20 (or 16·7%) had less than the maximum available number of letters arranged in a row. In confirmation of earlier data, a bias towards the high row was shown by the distribution of rows used: H, 53%; M, 41%; L, 6%. Within each subject's records, there was a high degree of consistency in the rows chosen on successive trials in any condition; the preponderance of the most frequently chosen row over the other two was in the ratio 5·25 : 1.

By contrast, behaviour in the whole/partial condition was highly variable. Although by chance exactly the same number of letters was attempted, the spatial distribution of these letters showed little consistency. Of the 120 trials recorded, only 41 (or 34·2%) showed horizontal order, while 31 (or 25·8%) were clearly aligned vertically; the remaining 40% of trials showed various mixed strategies. Within subjects, the most often repeated configurations only composed 22·5% of the trials.

Discussion

What might be called the "aniconic" position appears to remain tenable after further exploration of partial reporting behaviour. It seems that subjects can take in three or four letters at the position where they centre their gaze or attention, and a further one or two letters at closely adjacent positions. They cannot,

apparently, refer back to a passive image in order to extract cued information from previously non-attended rows. Depending upon the circumstances, they may attempt to maximize their scores by fixating expected letter rows, middle rows, or vertical or mixed configurations. To the extent that they are uncertain what will be demanded of them, they will tend to be indecisive and variable in what is selected for attention and rehearsal, with some reduction in the efficiency of both immediate report and retention.

Viewing the partial report as a sampling technique seems unnecessary, if the number of letters which are registered do not exceed the memory span. The span for letters has been estimated at six or seven letters by Crannell and Parrish (1957), using conventional verbal methods. It has been suggested by Neisser (1967) that the effective span is reduced in the tachistoscopic experiment by the high rate of readout required from the fading icon, as it is by the computer-accelerated presentation of digits. However, this line of argument begs the question whether the icon exists, and leads to other inconsistencies; thus, a span of four letters is associated in the study cited with a presentation rate of ten per sec, so that with a Baxt-type erasure following a 50-msec exposure the subject should read half a letter, rather than the three observed by Sperling (1960). For what it is worth, the impression of subjects in these experiments has been, not that they could see more than was remembered, but that in the time available three or four letters were all that could be clearly seen.

It is not argued, of course, that perceptions terminate instantaneously as stimulation ceases; but fading times of the order of milliseconds suggested by the durations over which backward masking is effective are very different from the period of up to a second during which the visual image has been supposed to preserve information. There is little doubt that, once selected for attention, information may be still available in a quasi-visual form at considerably longer delays; it is known, for example, that in aiming at a target no longer visible, accuracy continues to decrease lawfully at delays of several seconds (Holding, 1968). However, there is a variety of explanations in terms of storage and rehearsal mechanisms, perhaps even taking the form of regenerated images, which may be appropriate to these cases.

Whether any new information may be considered available at a delay after input depends upon the possibility that attention may be unselective. Stored perceptual data may possibly be restructured or reinterpreted on the advent of delayed cues, but this process should be distinguished from the delayed extraction of new information. Such extraction clearly requires that the subject can pay sufficient attention to a complex input to preserve the information without an active or passive process of selection. On the basis of the present data, there are no grounds for separating attention from selection in this way.

References

Bryden, M. P. (1961). The role of post-exposure eye movements in tachistoscopic perception. *Canadian Journal of Psychology*, **15**, 220–25.

Crannell, C. W. and Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters and words. *Journal of Psychology*, **44**, 319–27.

HOLDING, D. H. (1968). Accuracy of delayed aiming responses. *Psychonomic Science*, **12**, 125–6.

HOLDING, D. H. (1970). Guessing behaviour and the Sperling store. *Quarterly Journal of Experimental Psychology*, **22**, 248–56.

NEISSER, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

PAGE, E. B. (1963). Ordered hypotheses for multiple treatments: a significance test for linear ranks. *Journal of the American Statistical Association*, **58**, 216–27.

SPERLING, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74, Whole No. 498.

Received 2 June 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 82–96

EFFECTS OF REPETITION OF DISPLAY COMPONENTS ON THE LATENCY OF MULTIPLE REPORTS OF CONGRUENCE

PHILIP H. K. SEYMOUR

Department of Psychology, The University, Dundee

Subjects gave grouped multiple reports of the congruence of each member of arrays of one, two or three word-shape or shape-shape pairs, and a measure was taken of the time elapsing between onset of the array and initiation of a multiple yes/no report. Double and triple reports were initiated less rapidly than single reports. Replicated arrays, involving repetition of a display pair, gave similar response times for double and triple reports, and were classified faster than non-replicated arrays requiring the same overt report. In the case of non-replicated arrays triple reports were initiated less rapidly than double reports. Both classes of array showed substantial effects for congruence, giving slow response times where all pairs in the array were incongruent or where the lefthand or first reported display was incongruent.

Introduction

Currently available models of "same"—"different" judgements rely to a considerable extent on the observation that times to indicate that two stimuli are *same* are sometimes slower than times to indicate that they are *different* (Bindra, Williams and Wise, 1965; Nickerson, 1969), and sometimes faster (Bindra, Donderi and Nishisato, 1968; Bamber, 1969). Bamber has shown that the time required for a judgement that an array of letters held in memory is *different* from an array displayed in a tachistoscope can be described by a serial self-terminating search model (Egeth, 1966). However, times for the judgement *same* were faster than predicted by this model, and Bamber concluded that his results implied the parallel operation of two judgemental processes, an *identity reporter* which gave a fast output for *same*, and a slower *serial processor* capable of computing the degree of similarity between two stimuli. Seymour (1969*a*) found that judgements of the congruence of a printed name and a shape also showed faster times for *same* pairs than for *different* pairs. If interpreted in Bamber's terms, this would imply that the "identity reporter" is sensitive to nominal equivalence as well as to physical identity (cf. Posner and Mitchell, 1967).

Bamber (1969) pointed out that differences in latency for *same* and *different* judgements might arise if either one or both of two assumptions were false: (1) the assumption of *processing isochronality*, which states that, for a given set of stimuli, the time to compute sameness equals the time to compute difference, and (2) the assumption of *efferent isochronality*, which states that the time required to indicate a judgement of sameness equals the time required to indicate a judgement of difference. In general, models such as those of Nickerson (1969) or Bamber

(1969) question the assumption of processing isochronality but accept the assumption of efferent isochronality. On the other hand, Briggs and Blaha (1969) varied memory load and display load in a shape matching task, and reported a difference between "yes" and "no" responses which was independent of memory load. They interpreted this result as questioning the assumption of efferent isochronality, suggesting that longer was required to select a "no" report than a "yes" report. In this formulation a distinction is made between a stage of response *selection* and response *execution* (Smith, 1968), and it is argued that the failure of the assumption of efferent isochronality arises at the stage of selection rather than in execution of the response used to indicate the outcome of a judgement.

Seymour (1969*b*, 1970) examined the effects of mapping the verbal report "no" to congruent word-shape displays and the report "yes" to incongruent displays. This reversal did not eliminate the congruence effect, since times for congruent pairs remained faster than times for incongruent pairs, and hence supported the conclusion that differences between "yes" and "no" reports did not arise at the terminal stage of response execution. On the other hand, reversal of the congruence-report allocation did increase the difficulty of the task, raising the number of errors and slowing the verbal reaction time (VRT). This implies that the outcome of a decision about congruence is represented in a form which is semantically related to the verbal categories "yes" and "no", so that "yes" is compatible with the implicit response *same* whereas "no" is not. In other words, an overt report on the congruence of two stimuli may depend on (1) the computation of congruence, and (2) the occurrence of a mediating event which may, for convenience, be termed implicit *affirmation* or *denial*. The question about efferent isochronality is then one about the possibility of differences in the durations mediating affirmation or denial.

This paper attempts to explore this question by arranging an experimental task which varies the extent of implicit activity required to set up an overt report. A number of studies are now available which have examined rates of comparing items held in memory against items on a display (Sternberg, 1967; Briggs and Blaha, 1969; Bamber, 1969) or rates of computing "same"-"different" judgements for letter pairs (Beller, 1970). These studies provide estimates of the time required to judge the congruence of more or less complex stimulus sets, but do not vary the number of reports which must be given. In general, the subject is required to give a single report which indicates a *same* or *different* response to the relationship between two arrays, so that the RT includes perceptual and computational events and a single operation of translation (affirmation or denial) mediating between judgemental activity and a report. It seems likely that a version of the multiple comparison task in which subjects give grouped reports of the accumulated outcomes of a series of judgements of congruence will vary the extent of implicit affirmation or denial underlying an overt response.

This possibility was initially explored in a pilot study in which subjects responded to arrays of three word-shape pairs (the word *square* or *circle* printed inside a *SQUARE* or *CIRCLE*) by reporting on the congruence of the left-hand pair (single report), the left-hand and centre pairs (double report), or all three pairs

(triple report). The instruction was to classify congruent pairs as "yes" and incongruent pairs as "no", and to group the multiple reports as closely as possible. The pilot study indicated (1) that subjects accepted and conformed to the instruction to give *closely grouped multiple reports*, and (2) that the function relating delay of response initiation to number of reports was approximately linear. The pooled data of 12 subjects showed a relation between VRT and number of reports having a zero intercept of about 300 msec and a slope of about 400 msec. In other words, a report on the congruence of the left-hand word-shape pair was delayed by 400 msec if it was followed immediately by a report on the centre display, and by a further 400 msec if it was followed by reports on the centre and right-hand displays.

Further exploratory study of the multiple report task showed that an exception to the relationship between VRT and number of reports occurred where the word-shape displays forming the array were identical to one another. Examples of arrays involving *replication* of a word-shape display are shown in Figure 1, together

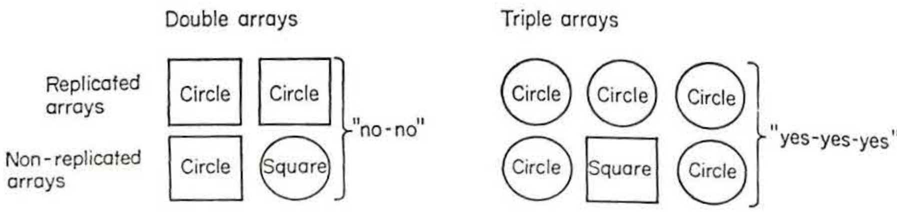

FIGURE 1. Examples of replicated and non-replicated double and triple word-shape arrays.

with examples of *non-replicated* arrays requiring the same overt report. Preliminary tests suggested that replicated arrays could be classified by reiterated "yes" or "no" reports considerably faster than non-replicated arrays, and that the slope of 400 msec per additional report observed in the pilot study was virtually eliminated when times for replicated arrays were examined. This effect appears analogous to Beller's (1970) finding that the time required to indicate the sameness of all members of a row of letters is independent of the number of letters in the array, over the range 2 to 8 letters, and also has features in common with Sanders's (1967) report that the spatial arrangement of signals on a display (S-S compatibility) influences the time required to organize a grouped key-pressing response.

The present experiment was designed to examine the distinction between replicated and non-replicated arrays more thoroughly, and in particular to investigate the hypothesis that the two classes of array differ in terms of the amount of mediating activity required to set up a report. More explicitly, the detection of replication across an array of displays of the type shown in Figure 1 may permit the subject to select a reiterated "yes" or "no" report on the basis of a single implicit response (affirmation or denial) to the congruence/incongruence of the array as a whole. In this case, times to classify replicated arrays should be more or less independent of the number of displays presented. On the other hand, if non-replicated arrays involve segmentation of the displays (Beller, 1970) and an implicit response to the congruence/incongruence of each display pair, times for these arrays should be slower than for replicated arrays, and should show the

relationship between number of reports and VRT found in the pilot study. Further, differences in VRT between replicated and non-replicated arrays, where array length and report length are held constant, should index covert labelling activity of the kind which has been discussed, and permit a comparison of the durations of implicit affirmation and denial.

Method

Subjects

Six volunteers acted as subjects. They were graduate or undergraduate members of the Faculty of Social Science at the University of Dundee. With one exception, they had had fairly extensive previous practice in word-shape matching and multiple report tasks.

Apparatus

The apparatus was set up to provide the following sequence of events: (1) After a verbal "Ready" from the experimenter a rear projection screen, measuring 8×8 in., was diffusely illuminated, and the bulb of a slide projector was pre-heated by switching the mains to it through a resistance. (2) Closure of a switch switched off the warning light and switched the mains abruptly to the projector bulb, and at the same time started a Venner millisecond stopclock. (3) The subject's vocal response closed the relay of a voice key, which stopped the clock and terminated the display.

Displays

The sets of displays used in the experiment were as follows: (1) *Single word-shape displays.* The set consisted of the four word-shape pairs produced by printing the word *square* or *circle* in lower case type (Letraset, Univers 55, 36 pt) inside a *SQUARE* or *CIRCLE.* The horizontal and vertical extent of the shapes when projected by the apparatus and viewed at a distance of about 60 in. was slightly less than 2° of visual angle. (2) *Double word-shape arrays.* The sets consisted of the 16 double arrays which are generated by printing *square* or *circle* inside a *SQUARE* or *CIRCLE* at each of two adjacent locations. Four of these arrays involved *replication* of a word-shape pair, and were classifiable by the reports "yes-yes" and "no-no". A further subset of four arrays did not involve replication, but were also classifiable by the reiterated reports "yes-yes" and "no-no". The remaining 8 arrays were classifiable by the reports "yes-no" and "no-yes", and involved either replication of the shape display or replication of the word display. (3) *Triple word-shape arrays.* A set of 16 triple arrays was selected from the 64 arrays generated by taking all possible sequences of three *SQUARES* or *CIRCLES* and printing the word *square* or *circle* at each location. Four of the 16 arrays involved replication of a word-shape display and were classifiable by the reiterated reports "yes-yes-yes" and "no-no-no". Four others were classifiable by the same reiterated reports, but the centre pair of the array was different from the other two pairs. The remaining arrays were classifiable by the reports "yes-no-yes" and "no-yes-no", and involved either replication of the shape display across the array, or replication of the word display. (4) In addition, sets of single and double shape-shape arrays were prepared. For the single displays a small *SQUARE* or *CIRCLE* was drawn inside an outline *SQUARE* or *CIRCLE.* For the double arrays, the 16 combinations described for the word-shape arrays were duplicated, the word being replaced by a small *SQUARE* or *CIRCLE* in each case.

All displays were photographed and mounted as 2×2 in. transparencies for projection by the apparatus.

Procedure

Testing on the word-shape arrays was carried out in three experimental sessions held on different days, and each involving 120 observations. Each session was organized as three

Table I

Means and standard deviations of VRTs in classification of single, double and triple word-shape arrays

		Replicated						Non-replicated				Partially replicated (shape)				Partially replicated (word)			
		Y	YY	YYY	N	NN	NNN	YY	YYY	NN	NNN	YN	YNY	NY	NYN	YN	YNY	NY	NYN
S1	$\bar{x}$	693	779	833	742	952	939	1138	1253	1413	1637	1258	1378	1072	1252	1295	1421	1333	1445
	S.D.	89	123	159	81	164	300	208	172	177	231	231	227	283	248	147	171	304	282
S2	$\bar{x}$	609	702	732	584	760	823	1073	1137	1246	1567	957	1132	932	1085	1021	1187	1014	1263
	S.D.	78	83	84	160	164	239	207	207	252	304	197	330	161	242	201	201	230	335
S3	$\bar{x}$	620	645	639	645	733	746	760	976	1103	1273	746	882	766	886	864	1071	870	1064
	S.D.	60	57	49	58	87	105	120	173	251	360	94	126	72	141	123	298	112	257
S4	$\bar{x}$	557	666	623	582	751	747	754	872	836	980	767	845	753	865	850	962	824	977
	S.D.	111	61	164	153	92	82	101	84	137	103	144	111	131	120	120	135	107	167
S5	$\bar{x}$	613	706	719	639	793	858	861	1050	1045	1296	929	1063	871	992	991	1204	977	1208
	S.D.	89	60	100	80	99	223	127	277	258	314	229	398	176	244	228	363	160	358
S6	$\bar{x}$	726	821	787	763	1022	1080	1092	1315	1385	1498	1087	1203	1119	1318	1363	1434	1163	1455
	S.D.	113	98	111	82	192	280	253	334	285	267	198	262	305	288	298	296	187	307
Mean		636	720	722	659	835	866	946	1101	1171	1375	957	1084	919	1066	1064	1213	1030	1235

blocks of 40 trials. Within a block, the 32 double and triple word-shape arrays occurred once each, and the four single arrays twice each. The orders of presentation of the displays were independently randomized for each subject at each block.

The 6 subjects were asked to return for a fourth session, in which latencies were recorded for single and double shape-shape arrays. This session was organized as five blocks of 20 presentations, with each of the 20 shape-shape arrays occurring once per block. Again, orders of presentation for each subject at each block were independently randomized.

Subjects were instructed to report on the congruence of each member of an array by giving a grouped multiple "yes/no" report, reading across the array from left to right. It was stressed that pauses between the elements of the report should be minimized, and that initiation of the report should be delayed until performance satisfied the experimenter's criteria on this point. An earlier pilot study confirmed (1) that the instruction to group reports closely could be obeyed consistently throughout an experimental session, and (2) that the stress placed on this requirement was important. In the pilot study subjects were able to gain some time if they were encouraged to go ahead with the first part of the report and give the remainder after a short pause. In general, a report of the form "yes . . . no-yes" was reliably faster than the forms "yes-no . . . yes" or "yes-no-yes". In the present study, a careful watch was kept on the closeness of the grouping of the multiple reports; where a failure of grouping was detected the subject was warned, and the observation was discarded and replaced.

Error responses were not treated as data, and slides associated with errors were reinserted later in the random series. Hence, all observations accepted for subsequent analysis were accurate reports which satisfied the grouping requirement imposed by the experimenter.

Results

Table I shows the means and standard deviations of the distributions of VRTs obtained in classification of the word-shape arrays (Sessions 1–3). Each mean is based on 18 observations, apart from means for single displays, which are based on 36 observations. The data have been classified in terms of class of array (replicated, non-replicated, or partially replicated by shape or word) and the overt report required. Table II gives the mean per cent error rates for replicated, partially replicated and non-replicated arrays, averaged over subjects. The error scores indicate that high frequencies of error occurred when reiterated reports were mapped to non-replicated arrays, but not when these same reports were

TABLE II

Mean per cent error rates for reports of congruence of word-shape and shape-shape pairs in single, double and triple arrays

	Single		Double		Triple
	Word-shape arrays	Shape-shape arrays	Word-shape arrays	Shape-shape arrays	Word-shape arrays
Replicated arrays (reiterated reports)	4·40	0·85	2·78	2·50	3·28
Partially replicated arrays (alternated report)			8·76	10·00	7·41
Non-replicated arrays (reiterated report)			20·36	10·00	16·61

mapped to replicated arrays. Mapping of alternated reports to partially replicated arrays was associated with intermediate error rates. The individual data for Session 4, involving classification of single and double shape-shape arrays, are given in Table III. Error rates for this session have been included in Table II. In order to facilitate discussion of these results, the means for the various sub-conditions of the experiment have been averaged across subjects, and are displayed in Figure 2.

TABLE III

Means and standard deviations of VRTs in classification of single and double shape-shape arrays

		Replicated				Non-replicated		Partially replicated			
								interior shape		exterior shape	
		Y	N	YY	NN	YY	NN	YN	NY	YN	NY
S1	$\bar{x}$	643	645	671	881	885	1047	809	1040	843	947
	S.D.	93	75	60	138	54	148	64	186	83	149
S2	$\bar{x}$	577	592	623	722	700	791	672	834	686	757
	S.D.	68	55	47	65	99	103	91	111	85	108
S3	$\bar{x}$	537	600	586	814	825	982	877	970	848	958
	S.D.	59	55	55	106	97	51	89	87	70	50
S4	$\bar{x}$	506	552	561	681	676	736	699	763	691	745
	S.D.	19	31	42	43	52	75	76	47	56	77
S5	$\bar{x}$	661	702	651	962	963	1177	971	1231	940	1126
	S.D.	73	81	55	180	117	102	156	217	97	150
S6	$\bar{x}$	566	622	635	837	1015	1201	938	1104	1103	1133
	S.D.	46	85	69	136	88	123	85	192	306	125
Mean		582	619	621	816	844	989	828	990	852	944

Replicated arrays

The group data for the replicated arrays are summarized as the two lower curves in Figure 2. The data for word-shape arrays will be considered first. As a preliminary, a comparison was made between the times for replications of *circle/CIRCLE* and *square/SQUARE*, since the display *circle/CIRCLE* had been shown to be associated with fast response times in a previous study (Seymour, 1969*a*). In this study, no difference was found between these two displays ($F = 1{\cdot}21$, $df = 1{,}5$). The main analysis of the times for the replicated arrays examined the factors of number of reports (1, 2 or 3) and congruence of the array within a two factor repeated measures design. This analysis showed an effect for number of reports ($F = 49{\cdot}52$, $df = 2{,}10$, $P < 0{\cdot}001$), an effect for congruence ($F = 24{\cdot}47$, $df = 1{,}5$, $P < 0{\cdot}01$), and an interaction between these factors ($F = 15{\cdot}42$, $df = 2{,}10$, $P < 0{\cdot}001$). The effect for number of reports occurred because single reports were faster than either double or triple reports. The congruence effect

occurred because times were faster for congruent arrays than for incongruent arrays. This effect appeared for all subjects under the multiple report conditions, and for 5 of the 6 subjects under the single report condition. However, the congruence effect was much smaller under the single report condition, and this is reflected by the significant congruence × reports interaction.

An analysis of single and double reports of shape-shape arrays gave essentially

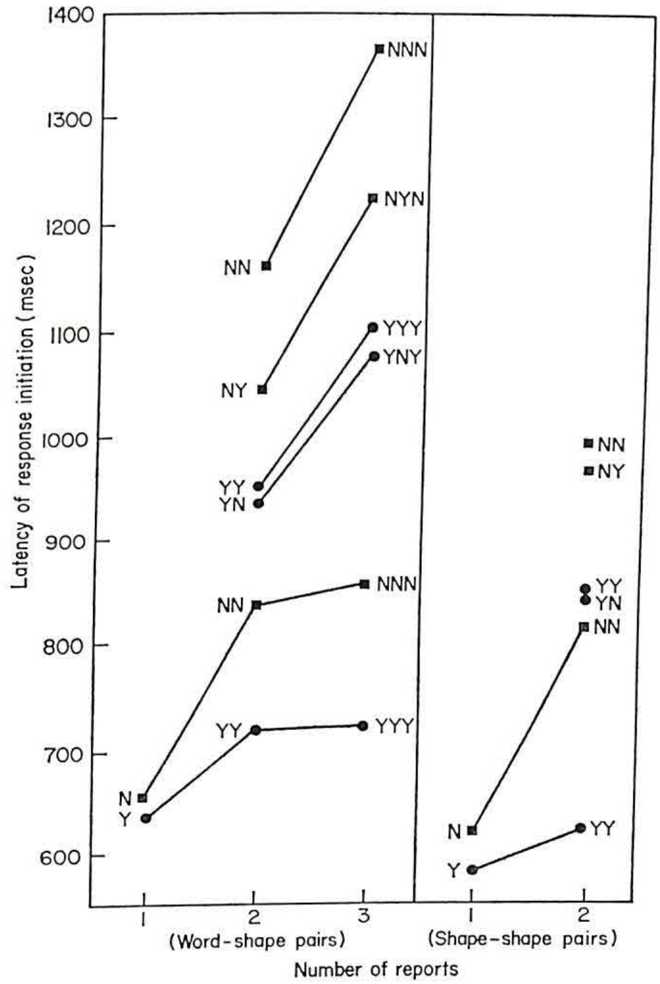

FIGURE 2. Mean latencies of response initiation for single, double and triple reports of congruence.

similar results. Congruent displays were classified faster than incongruent displays ($F = 43{\cdot}73$, $df = 1,5$, $P < 0{\cdot}01$), single displays were classified faster than double arrays ($F = 73{\cdot}01$, $df = 1,5$, $P < 0{\cdot}001$), and there was an interaction between these factors ($F = 26{\cdot}62$, $df = 1,5$, $P < 0{\cdot}01$). In summary, all subjects showed an effect for congruence, but this effect was much greater for the double arrays.

Non-replicated arrays

According to the hypothesis outlined earlier non-replicated arrays should be mapped to a reiterated report less rapidly than replicated arrays. Inspection of the individual data given in Table I, and the summary given in Figure 2, will confirm that this occurred. An analysis of variance was carried out on the mean response times for the double arrays, in which the factors of array type (*replicated vs. non-replicated*) and report ("yes-yes" vs. "no-no") were examined within a two-factor repeated measures design. This analysis showed a significant effect for array type ($F = 27{\cdot}87$, $df = 1{,}5$, $P < 0{\cdot}01$), and a significant effect for reports ($F = 32{\cdot}52$, $df = 1, 5$, $P < 0{\cdot}01$). The interaction between array type and reports was also significant ($F = 10{\cdot}74$, $df = 1{,}5$, $P < 0{\cdot}05$). This interaction reflects a tendency for the difference between replicated and non-replicated arrays to be greater for the report "no-no" than for the report "yes-yes". The differences involved have been summarized in Table IV, which shows that $YY_{nr} - YY_r = 227$ msec, but that $NN_{nr} - NN_r = 336$ msec. The individual results for all 6 subjects were consistent with this summary, in that all subjects showed slower times for non-replicated arrays, and all but one (S4) showed a larger effect of array type for incongruent arrays.

TABLE IV

Estimates of central translation time (msec) obtained by subtraction of times for congruent and incongruent replicated arrays from times for partially or non-replicated arrays

	Double arrays word-shape displays	shape-shape displays	Triple arrays	word-shape displays
YY_{nr}–YY_r	227	223	YYY_{nr}–YYY_r	378
YN_{pr}–YY_r	218	219	YNY_{pr}–YYY_r	353
NN_{nr}–NN_r	336	173	NNN_{nr}–NNN_r	510
NY_{pr}–NN_r	212	151	NYN_{pr}–NNN_r	359

Subscript r = replicated, pr = partially replicated, and nr = non-replicated.

The data for shape-shape arrays were similar to the results obtained for the word-shape arrays. All subjects recorded slower times for non-replicated arrays than for replicated arrays ($F = 18{\cdot}62$, $df = 1{,}5$, $P < 0{\cdot}01$), and slower times for the report "no-no" than for the report "yes-yes" ($F = 39{\cdot}55$, $df = 1{,}5$, $P < 0{\cdot}01$). However, the interaction between these factors ($F = 13{\cdot}20$, $df = 1{,}5$, $P < 0{\cdot}025$) reflected a tendency for the difference between replicated and non-replicated arrays to be *smaller* for incongruent arrays than for congruent arrays. This can be checked in Table IV, where $YY_{nr} - YY_r = 223$ msec, but $NN_{nr} - NN_r = 173$ msec.

These results are somewhat equivocal with respect to the question about efferent isochronality which was raised earlier. The data for both word-shape and shape-shape comparisons agree in showing faster responses to congruent replicated arrays than to incongruent arrays. This congruence effect is also

maintained in response times for the non-replicated arrays. On the assumption that the additional time required to classify a non-replicated array represents the duration of an implicit response to the second display, one can ask whether the delay is smaller for a congruent second display than for an incongruent one. For the word-shape arrays 5 of the 6 subjects show larger differences between replicated and non-replicated arrays for "no-no" reports than for "yes-yes" reports. On the other hand, for shape-shape arrays, all 6 subjects show larger differences for "yes-yes" reports.

If, in the case of non-replicated arrays, an implicit response must occur to each of the displays forming the array, triple non-replicated arrays should give evidence of delays relative to triple replicated arrays equivalent to the duration of two successive covert mediating events. Table IV gives the results of subtraction of times for non-replicated arrays from times for replicated arrays for "yes-yes-yes" and "no-no-no" reports. These results agree with the finding for the double word-shape arrays in showing a larger difference for the incongruent arrays. An analysis of variance on times for the triple reports confirmed the significance of the effects for congruence ($F = 78{\cdot}5$, $df = 1{,}5$, $P < 0{\cdot}001$) and type of array ($F = 71{\cdot}45$, $df = 1{,}5$, $P < 0{\cdot}001$), but not of the interaction between these factors ($F = 2{\cdot}75$, $df = 1{,}5$). Inspection of the individual results indicates that the difference for "no-no-no" reports is quite considerably greater than for "yes-yes-yes" reports for 4 of the 6 subjects; one subject, S4, shows very little difference, and for S6 there is a substantial difference in the opposite direction.

All subjects showed a slowing of response times in moving from double to triple non-replicated arrays, the mean delay being 165 msec for congruent arrays and 204 msec for incongruent arrays. An analysis of variance indicated that the delay was significant ($F = 337{\cdot}14$, $df = 1{,}5$, $P < 0{\cdot}001$), but that there was no interaction between number of reports and congruence.

Partially replicated arrays

The remaining observations involved arrays in which either the word or the shape was replicated across the array, but the other component varied. Preliminary analyses were carried out to determine whether it was important which component of the array was replicated: analyses of variance failed to show a reliable effect for this factor either for double arrays ($F = 6{\cdot}39$, $df = 1{,}5$) or for triple arrays ($F = 1{\cdot}27$, $df = 1{,}5$). In the subsequent analyses times for the two types of partially replicated array were combined.

If a double replicated congruent array is altered by changing the value of the word or shape component of the second display, the correct report becomes "yes-no", and a delay of response initiation occurs which is closely comparable to that observed for non-replicated congruent arrays, i.e. 218 msec. This result is well supported by the data for the shape-shape arrays, where the mean delay relative to replicated arrays was 219 msec. Replicated arrays classifiable by a "no-no" report show comparable results if the value of one component of the second display is altered, since "no-yes" reports were 212 msec slower than reports of replicated incongruent arrays for the word-shape displays, and 151 msec slower for the shape-shape arrays.

These results can be considered from the viewpoint of the assumption of efferent isochronality. If the assumption does not hold, so that, for example, implicit affirmation is an event of shorter duration than implicit denial, a "yes-no" report might be expected to have a slower latency than a "yes-yes" report for a non-replicated array; similarly, a "no-yes" report should be faster than a "no-no" report of a non-replicated array. The results for the shape-shape arrays provide little support for such an account. An analysis of variance which examined the factors of first report ("yes" vs. "no") and report type (*reiterated* vs. *alternated*) confirmed that there was an effect for first report ($F = 38{\cdot}17$, $df = 1{,}5$, $P < 0{\cdot}01$), but showed no effect for *reiteration* vs. *altenation* of the report, and no interaction between these factors.

The same question was examined for the word-shape arrays in an analysis which also considered the transition from double to triple reports. The factors of report length (one or two reports) and report type (*reiterated* vs. *alternated*) were examined for non-replicated and partially replicated arrays having a congruent first display. This analysis confirmed that triple reports were slower than double reports ($F = 130{\cdot}42$, $df = 1{,}5$, $P < 0{\cdot}001$), but showed no effect for type of report ($F = 1{\cdot}0$, $df = 1{,}5$) and no interaction between this factor and report length. This result agrees with the finding for the shape-shape arrays, and is equally consistent with the assumption of efferent isochronality.

The prediction for arrays in which the first display is incongruent is that "no-yes" and "no-yes-no" reports should be faster than "no-no" and "no-no-no" reports of non-replicated arrays. An analysis of variance of these response times confirmed the effect of length of report ($F = 73{\cdot}46$, $df = 1{,}5$, $P < 0{\cdot}001$), and also showed an effect for report type ($F = 8{\cdot}81$, $df = 1{,}5$, $P < 0{\cdot}05$), but no interaction with the number of reports. This result differs from that obtained for the shape-shape arrays, in that it shows reduction in response times resulting from inclusion of a congruent display within an otherwise incongruent array. The effect appears in the data of all subjects except S4.

Discussion

For a single word-shape or shape-shape display the VRT for saying "yes" is slightly faster than the VRT for saying "no". This effect is present in the data for all 6 subjects in the shape-shape comparison task, and for 5 of the 6 subjects in the word-shape comparison task, although the size of the difference was smaller in this experiment than in previous studies (Seymour, 1969*a*,*b*). As stated earlier, this effect can be attributed to failures of one or both of the assumptions of processing or efferent isochronality.

The multiple report task was introduced with the aim of developing an index of the covert activity involved in translating from a decision about congruence to a report. In general, the index used was delay of response initiation where array length and report length are held constant, but the internal structure of the array is varied. The usefulness of such an index in relation to questions about efferent isochronality is, however, dependent on the possibility of giving an adequate account of the subject's behaviour in the multiple report task.

Sanders (1967) has reported that the latency of a grouped response to two

signals is influenced by S-R compatibility (the relative arrangement of display and response units) and S-S compatibility (the arrangement of signals on the display). It appears likely that both of these concepts are relevant to a discussion of the multiple report task. Pilot work with multiple report tasks suggested that subjects use the *configuration* of an array in planning the general form of a report, since errors were often the complement of the correct report. In the present experiment, error frequencies (see Table II) were consistent with the view that variations in S-R compatibility occurred when different classes of report (reiterated vs. alternated) were mapped to different classes of array (replicated vs. non-replicated). Errors were fewest where report structure matched array structure (reiterated reports of replicated arrays) and increased when alternated reports were mapped to partially replicated arrays, or reiterated reports were mapped to non-replicated arrays. However, this line of explanation is not adequate to handle the latency data, since it does not account for the congruence effects which were observed.

Since the variations in response time which occurred were correlated with changes in the internal structure of the arrays it may be more profitable to consider the task in terms of S-S compatibility. In general, S-S compatibility was manipulated in this experiment by repetition of display elements horizontally across an array (replication) or vertically within a display pair (congruence). Replication can be discussed in terms of two rows, defining the interior and exterior location of the displays, in which case the term *replication* refers to physical duplication of a word or shape within one or both rows of the array. Vertical repetition, or congruence, refers to physical duplication (shape-shape arrays) or replacement by a conceptually equivalent signal (word-shape arrays). Display and response repetition effects have previously been studied in the context of serial choice-response tasks (Rabbitt, 1968; Eichelman, 1970). However, Entus and Bindra (1970) have argued that facilitation for *identical transitions* (Rabbitt, 1968) in serial choice tasks and for *same* stimulus sequences in "same"-"different" judgemental tasks reflects the operation of a common process. Eichelman (1970) has also pointed out similarities between *identical* and *equivalent transitions* in serial choice tasks and *physical* and *name identity* matches in the "same"-"different" judgemental task (Posner, Boies, Eichelman and Taylor, 1969).

The data obtained in the present study suggest that the multiple report task is comparable to the serial choice-response task with respect to repetition effects. Replicated arrays (identical transitions) were classified faster than non-replicated arrays requiring the same report (equivalent transitions). Reiteration of the vocalisation "yes" or "no" did not show a significant advantage over alternated reports (new transitions). However, the arrays classified by alternated reports in this experiment cannot be regarded as analogous to those occurring in a serial choice-response task, since they always involved replication of one display component across the array.

These comments suggest that it may be possible to describe a subject's performance in the multiple report situation in terms of his ability to detect and make use of horizontal and vertical repetition of display elements. Maximum facilitation occurs where there is duplication both horizontally and vertically, i.e. in the

case of replicated congruent shape-shape arrays. These arrays were classified as "yes-yes" about 100 msec faster than the equivalent word-shape arrays. This effect, which is shown by all subjects, is consistent with the distinction between name and identity matches observed by Posner and Mitchell (1967). Detection of horizontal replication appears to be impaired for both word-shape and shape-shape arrays if the displays are incongruent, since $NN_r - YY_r = 195$ msec (shape-shape arrays) and 115 msec (word-shape arrays), and $NNN_r - YYY_r = 143$ msec. This substantial effect for congruence results from elimination of vertical duplication or equivalence from the array, although horizontal repetition within the rows of the array is still influencing performance, since replicated arrays are classified consistently faster than non-replicated arrays.

There are perhaps three possible approaches to the problem of explaining this congruence effect: (1) it may be a response effect, arising because "no-no" is more difficult to articulate than "yes-yes". This is an improbable explanation, since differences between the two reports do not appear in simple VRT tasks. Also, the latencies of the two reports can be very similar in some multiple report situations, e.g. where the subjects report on the relationship between a single word and an adjacent array of shapes. (2) It may be an S-R compatibility effect, occurring at the stage of response selection because the report "no" is mapped to a display which involves sameness (horizontal replication) as well as difference (incongruence). (3) The effect may be perceptual, and related to the difficulty of detecting horizontal replication in incongruent arrays. Both replicated and non-replicated arrays show a tendency for VRT to depend on the congruence of the "first" or left-hand display. If the subject first checks the identities of the components of this display and then searches for replication across the array, he can do this in terms of successive inputs to a single analyser for congruent arrays, but must use two analysers for incongruent arrays. (This assumes that *square* and *SQUARE* are classified in terms of the activity of a single analyser.) It is difficult to choose between explanations in terms of S-R compatibility and delay in detection of replication, except to point out that the perceptual explanation has slightly greater generality since delays attributable to incongruence of the left-hand display also appear in the data for non-replicated arrays which do not involve the same degree of horizontal repetition as the replicated arrays.

The perceptual explanation identifies a search for horizontal replication as part of the subject's initial response to onset of the array. The data for replicated and non-replicated arrays can be quite well represented by assuming that the outcome of such a search is used to select between options (1) to respond to the relationship between a row of duplicated words and a row of duplicated shapes, or (2) to segment the array, and respond serially to each word-shape pair. Where replication is detected, response times are fast, few errors occur, and triple reports have similar latencies to double reports. Sanders and Keuss (1969) also reported that grouped responses were slower than single responses, but that the number of responses to be grouped did not influence RT (over the range 2–4 S-R combinations). Where replication is not detected, an overall slowing in VRT occurs which is independent of array length and report characteristics. For example, $YY_{nr} - YY_r = 227$ msec (word-shape arrays) and 223 msec (shape-shape arrays). This

result is at least consistent with the view that the task involves the organization of two independent "yes" reports in the case of non-replicated arrays, but only a single (reiterated) report in the case of the replicated arrays. At the same time, the effect can be described in terms of S-R compatibility, if it is argued that delays occur if a reiterated report must be mapped to a non-replicated configuration.

The delay which occurs as a consequence of increasing array and report length ($YYY_{nr} - YY_{nr} = 165$ msec) is also consistent with an account expressed in terms of serial responses to each additional word-shape display. This is not a simple response effect, since it does not occur when the same reports are given to replicated arrays. Comparable delays occur for the other response categories ($NNN_{nr} - NN_{nr} = 204$ msec; $YNY_{pr} - YN_{pr} = 137$ msec; $NYN_{pr} - NY_{pr} = 178$ msec). These delays are smaller than those observed in the pilot study described earlier, where the mean for triple reports was about 400 msec slower than the mean for double reports. This discrepancy may have arisen because subjects were more practised in the present study, or because of redundancy in the triple reports. (In the pilot study the third element of a triple report was not predictable given the outcomes of the first two judgements.)

The relevance of the comparisons given in Table IV to questions about the efferent isochronality of "same" and "different" judgements depends on the adequacy of the view that serial affirmation or denial of the congruence of each display underlies the reports given to non-replicated arrays. In general, comparisons involving arrays having a congruent left-hand display are consistent with the assumption of efferent isochronality, since $(YY_{nr} - YY_r) \not< (YN_{pr} - YY_r)$ and $(YY_{nr} - YYY_r) \not< (YNY_{pr} - YYY_r)$. Comparisons for arrays having an incongruent left-hand display were, on the other hand, indicative of delays correlated with handling of successive incongruent displays: for example $(NN_{nr} - NN_r) > (NY_{pr} - NN_r)$, $(NNN_{nr} - NNN_r) > (NYN_{pr} - NNN_r)$. These effects occur because the differences between VRTs for replicated and non-replicated arrays are greater for incongruent arrays $(NN_{nr} - NN_r)$ than for congruent arrays $(YY_{nr} - YY_r)$. This does not occur for alternated reports, since $(YY_{nr} - YY_r)$ is approximately equivalent to $(NY_{pr} - NN_r)$. Thus, evidence favouring the view that a covert response to an incongruent display (denial) is slower than a response to a congruent display (affirmation) appears only in the case of word-shape arrays having an incongruent left-hand display.

Apart from the NN_{nr} and NNN_{nr} reports of word-shape arrays the results of this experiment can be quite well described in terms of the factors of *replication* vs. *non-replication* and *congruence* vs. *incongruence* of the left-hand or first display. Elimination of horizontal repetition from the array results in a delay in initiation of a double report, and a further delay for a triple report. This effect can be interpreted as suggesting that the non-replicated arrays are segmented, and that serial covert responses occur to each display. If the left-hand display is incongruent reports on replicated and non-replicated arrays are delayed by 100 msec or more ($NN_r - YY_r = 115$ msec; $NNN_r - YYY_r = 143$ msec; $NY_{pr} - YN_{pr} = 109$ msec; $NYN_{pr} - YNY_{pr} = 150$ msec). The most reliable comparison here is between the NY and YN reports, since in this case the displays and report components are identical apart from order of arrangement. The processing

requirements for these two reports appear to be the same, since both involve the classification of one congruent and one incongruent display. The difference between them must, therefore, be related to the order in which the displays are handled. A similar effect occurred for the shape-shape arrays ($NY_{pr} - YN_{pr} =$ 128 msec). The implication, as suggested earlier, is that an input to two analysers at the outset of a scan across an array reduces the capacity of the visual system to handle immediately subsequent discriminations.

Thanks are due to the subjects who submitted themselves to this experiment, and to Maria Attivissimo, Janet Carter and Elizabeth Thomson who assisted in collection and analysis of the data.

References

BAMBER, D. (1969). Reaction times and error rates for "same"-"different" judgements of multidimensional stimuli. *Perception and Psychophysics*, **6**, 169–74.

BELLER, H. K. (1970). Parallel and serial stages in matching. *Journal of Experimental Psychology*, **84**, 213–9.

BINDRA, D., WILLIAMS, J. A. and WISE, J. S. (1965). Judgements of sameness and difference: Experiments on decision time. *Science, New York*, **150**, 1625–7.

BRIGGS, G. E. and BLAHA, J. (1969). Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology*, **79**, 395–402.

EICHELMAN, W. H. (1970). Stimulus and response repetition effects for naming letters at two response-stimulus intervals. *Perception and Psychophysics*, **7**, 94–6.

EGETH, H. E. (1966). Parallel vs. serial processes in multi-dimensional stimulus discrimination. *Perception and Psychophysics*, **1**, 245–52.

ENTUS, A. and BINDRA, D. (1970). Common features of "repetition" and "same-different" effects in reaction time experiments. *Perception and Psychophysics*, **7**, 143–8.

NICKERSON, R. S. (1969). "Same"-"different" response times: A model and a preliminary test. *Acta Psychologica*, **30**, 257–75.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H. and TAYLOR, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychological Monograph*, **79** (No. 1, Pt. 2).

POSNER, M. I. and MITCHELL, R. G. (1967). Chronometric analysis of classification. *Psychological Review*, **74**, 392–409.

RABBITT, P. M. A. (1968). Repetition effects and signal classification strategies in serial choice-response tasks. *Quarterly Journal of Experimental Psychology*, **20**, 232–40.

SANDERS, A. F. (1967). The effect of compatibility on grouping successively presented signals. *Acta Psychologica*, **26**, 373–82.

SANDERS, A. F. and KEUSS, P. J. G. (1969). Grouping and refractoriness in multiple selective responses. *Acta Psychologica*, **30**, 177–94.

SEYMOUR, P. H. K. (1969*a*). Matching latencies for word-shape pairs. *Quarterly Journal of Experimental Psychology*, **21**, 312–21.

SEYMOUR, P. H. K. (1969*b*). Response latencies in classification of word-shape pairs. *British Journal of Psychology*, **60**, 443–51.

SEYMOUR, P. H. K. (1970). Conceptual uncertainty and the latency of judgements of the congruence of word-shape pairs. *Acta Psychologica* (in press).

SMITH, E. E. (1968). Choice reaction time: an analysis of the major theoretical positions *Psychological Bulletin*, **69**, 77–110.

STERNBERG, S. (1967). Scanning a persisting visual image vs. a memorized list. Paper read at Eastern Psychological Association, Boston, U.S.A.

Received 8 *May* 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 97–106

RETINAL IMAGE OR PERCEIVED FEATURES AS DETERMINANTS OF ERROR IN GEOMETRIC ILLUSIONS?

R. I. HORRELL

Department of Psychology, University of Hull

An examination of the implications of the theory of geometric illusions proposed by Gregory (1963) leads to the prediction that the illusion error should be determined by the retinal image projected by the figure rather than by its perceived characteristics where these differ. This prediction is tested in two experiments, using a version of the Zöllner illusion. In the first experiment, distance is used to manipulate the retinal image characteristics and, in the second, slant. In both cases, error is found to follow changes in the retinal image rather than being held relatively constant by the perceived size and shape. It is suggested that the effect of slant and distance on illusions, as reported by others, can be attributed to the change in retinal image characteristics involved in such operations.

Introduction

Interest in the geometric illusions has recently been revived by Tausch (1954) who has suggested that they are related to normal constancy. This theory has been developed and extended by Gregory (1963) whose version has come in for considerable critical comment. The full implications of the theory have not always been clearly acknowledged by the critics and a brief reconsideration of the relations involved in it may clarify some apparent misconceptions and should indicate the relevance of the hypothesis here put to test.

The theory proposes two different mechanisms for setting constancy:

(i) The so-called "secondary mechanism". This is the embodiment of the invariance hypothesis and is the one most readily appreciated. In this, distance and depth are perceived on the basis of depth cues and then the perceived size and shape of a given retinal image is determined by the depth seen.

(ii) The "primary mechanism". In this, certain cues which normally indicate distance determine perceived size and shape directly: they do this quite independently of any effect they may have on perceived depth. For example, a given linear perspective configuration will invariably scale perceived size and shape in the same way through this channel, whatever the perceived depth in the situation.

The block diagram (Fig. 1) indicates the relationship between the two mechanisms. The two share some of the same operating cues. But they constitute two separate processing channels and may use the cues in independent ways. ("Channel" is used here in a sense that does not necessarily imply discrete, identifiable

physical embodiments but simply conceptually distinct processes.) The illusions themselves show how the same cue may not have the same effect in the two channels. It is supposed, for example, that primary scaling is operated by the linear perspective features of a Müller-Lyer figure. But they do not cause any

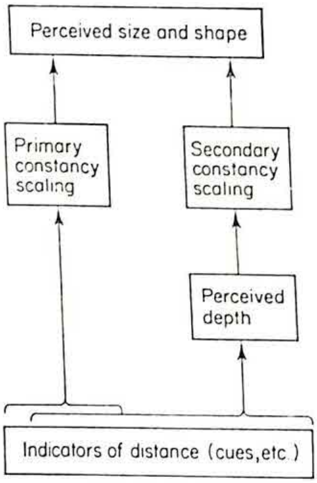

FIGURE 1

depth to be perceived because it is over-ruled by the textural features and lack of binocular disparity indicating a flat surface; and if there is no perceived depth, there is no secondary scaling. The two mechanisms do not even use all the same cues. Although a cue like linear perspective seems to be used by both, texture gradients can cause perceived depth but not illusions (Brown and Houssiadas, 1965).

Brown and Houssiadas (1965) predicted from this theory that there should be a tendency to see depth in illusions on the flat page and that the perceived error should be in a direction consistent with this depth. But their argument does not appear to recognize that, being two independent and discrete mechanisms, what happens in one does not necessarily happen in the other. And, though they themselves show that texture has no effect on the primary channel, it may be just this cue which is mainly responsible for nullifying the perception of depth in the secondary channel. Hamilton (1966) and Carlson (1966) have each correlated the tendency to see illusions with the extent of normal constancy across subjects. but, whereas the illusion error is determined by primary scaling, their constancy test situation probably involves secondary scaling; and their prediction that the correlation should be high seems to ignore the implication that there need be no relation between the relative efficiency of a person's primary and secondary mechanisms. Fisher (1968) has also failed to recognize the significance of the distinction between the two mechanisms. He used Müller-Lyer figures with parallel fins. Contrary to his own expectations from the theory, he found (i) no significant illusory effects in these figures and (ii) no changes in apparent size with reversals in apparent depth. But these findings are entirely consonant with the theory: there is no converging or diverging linear perspective in the figures to set

primary scaling and the apparent depth in the figures derives presumably from the secondary mechanism.

In all these cases and others, authors appear to have confused the separate effects of the two mechanisms, correlating the output of one with that of the other. Of course, if primary scaling does operate as an adaptive constancy scaling process in real-life, the outputs of the two channels ought to be similar in general; but the whole point about an illusory situation is that there is a discrepancy between the two and the prediction certainly cannot be made that they should always correlate precisely.

In a more recent account of the theory, Gregory (1968) has emphasized that the secondary mechanism (which is now described as performing "depth hypothesis scaling") is not tied to retinal information in the simple manner shown in Figure 1 but that such data are used to identify stored information on typical depthful situations experienced in the past. In other words, a very important feature of secondary scaling involves the operation of certain internal memory systems. To this extent, representation of the secondary channel in Figure 1 is greatly oversimplified. However, the essential point, that there are two distinct constancy scaling processes and that they operate in parallel, remains. This must be so to explain the facts if geometric illusions are to be regarded as cases of misapplied constancy. If the "depth hypothesis" is switched (as in a reversal of apparent depth) in conditions of constant retinal stimulation, perceived size changes may occur which follow the changes in apparent depth (this is best illustrated by the reversing self-luminous skeleton cube—Gregory, 1963): this implies something like secondary scaling. And a process like primary scaling is needed to explain the fact that the illusions occur in the absence of perceived depth, as well as such facts as those described by Brown and Houssiadas (1965), Carlson (1966) and Hamilton (1966).

It is the essence of normal size and shape constancy that the perceived size and shape of objects remains constant while the retinal image varies. The retinal image shape is altered when the orientation of a surface to the line of sight is changed; and the projected size can be manipulated by altering its distance from the observer. But, in both cases, perfect secondary scaling would keep the perceived size and shape constant. The size and shape of the inducing features affect the nature and size of an illusion. It would seem to be a necessary prediction of the above theory that, when the angle of regard of an illusion figure or its distance is manipulated, the illusion error should change with the changing retinal image rather than be held constant by the constant perceived size and shape of the inducing features. If the perceived shape determined the error, it would imply that the illusion causing mechanism depends for its input on the output of normal constancy. The block diagram of Figure 1 would then have the primary mechanism in series with the secondary rather than in parallel with it. This would not then be the system that is postulated; nor would it make sense to call the primary mechanism a separate constancy scaling device if this were the case. As it stands, the system implies that the error must be determined by the retinal image features rather than the perceived characteristics where these differ.

The purpose of the two experiments here reported was to test this prediction

using, in one case, distance and, in the other, slant to manipulate the retinal image. Houssiadas and Brown (1963) have found that the error decreases in many illusions when they are presented at a slant and when their distance from the observer is increased. Marshall and Di Lollo (1963) have also shown an effect due to distance. The problem is to determine whether these effects can be attributed to the change in retinal image involved in these procedures. The distortion of two parallel lines by a herring-bone background (a form of the Zöllner illusion) was here used to ascertain this.

Experiment I

In this experiment, the effect of increasing the distance of the figure (decreasing the projected but maintaining the object size) was compared with that of decreasing the size of the figure at the original distance.

Materials and Methods

Apparatus

The apparatus consisted of a wooden framework supporting a display surface (Fig. 2). The latter was pivoted about a horizontal axis so that displays could be presented at an

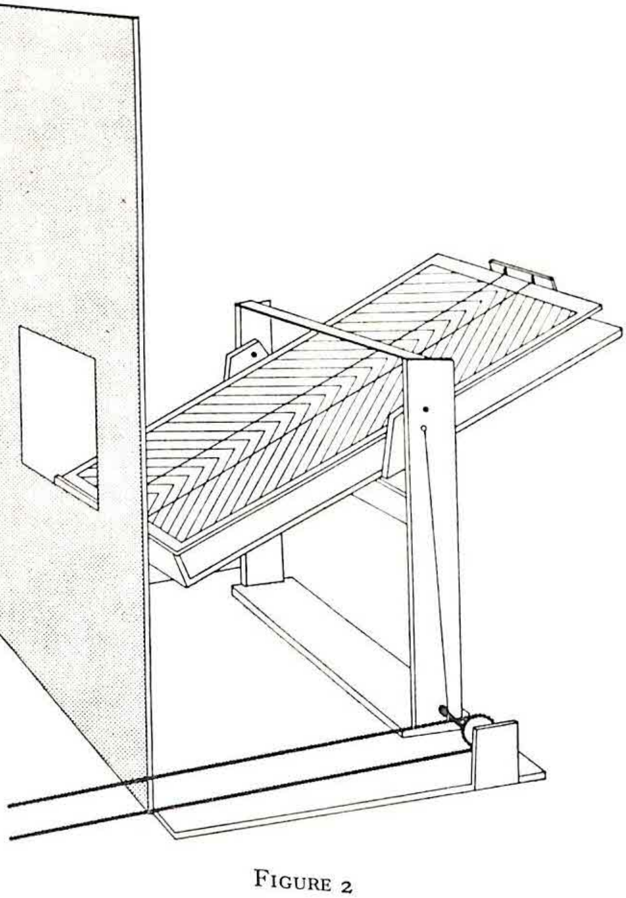

FIGURE 2

angle to the fronto-parallel plane, as shown in the figure: only vertical orientations were used in this experiment. Two taut black strings, in front of and in the same plane as the

display, were fixed at the top end but could be moved apart or together at the bottom and by means of a mechanical system which the subject could control by manipulating a continuous chain drive. The net result was that he could adjust the two strings to appear parallel. The actual separation of the two strings at the bottom could be read off a linear scale invisible to the subject.

The displays were two herring-bone figures with lines at 55° to each other and displayed so that the V shapes were inverted. One had its 0·6 mm thick lines separated by 3·0 mm; the dimensions of the other figure were one-third of these. A screen 18·0 cm in front of the display restricted the subject's view to the display. The rectangular aperture in this screen was 15·0 × 10·0 cm for the larger figure and 5·0 × 3·33 cm for the smaller. A narrower gauge string, approximately one-third of the diameter of the other, was used for the smaller display and the top ends of the strings were separated by 1·67 cm instead of 5·00 cm.

The subject's head was kept in the correct position within limits of ±2 cm by a face mask. This allowed binocular viewing and small head movements to maintain maximal depth cue availability. The display was lit by the same diffuse room illumination throughout and those subjects who normally wore spectacles did so.

Procedure

Six male and six female students were used as subjects. They were given a short preliminary demonstration of how to control the apparatus and were then asked to set the two vertical lines to appear parallel under each of three conditions:

(A) The larger display at 85 cm distance from the subject.

(B) The larger display at 255 cm [three times as far as in A]. In fact, the apparatus was fixed and S moved back to the required distance.

(C) The smaller display (a third of the size of the larger) at 85 cm.

Order of presentation of the conditions was controlled in two ways. First, two subjects were assigned to each of the six permutations of order. Second, each subject made two settings for each condition going through them in the allotted order and two more for each coming back through in the reverse order. Of these four settings made by each subject on each condition, two were made with the lines initially converging towards the bottom and two with them initially diverging.

Results

The raw score recorded for each reading was the separation of the two strings at the bottom. This was converted into a measure of the extent to which the lines were out of parallel by subtracting from it the separation of the two strings at the top. Means for each condition were computed and converted to the angle the two "parallel" lines made to each other from the geometry of the display. These two statistics are given in Table I.

TABLE I

Results for Experiment I

Condition	Mean error (cm)	Mean angle of lines to each other
A. Standard	1·70	3° 30′
B. Distant	1·09	2° 16′
C. Small	0·96	2° 1′

Means for each S on each condition were used for the statistical analyses (Student's t for correlated scores). The mean error for condition A is significantly greater than those for either condition B ($t = 5{\cdot}38$; $P < 0{\cdot}001$ with a two tail assumption) or condition C ($t = 11{\cdot}5$; $P < 0{\cdot}001$). The difference between conditions B and C is not significant ($t = 0{\cdot}94$). Thus the error in the distant condition is essentially the same as the nearer condition having the same retinal image rather than the same object size.

Experiment II

Here, the effect of slanting the figure (which increases the angle projected by the inverted Vs) was compared with that of a figure in the fronto-parallel plane projecting the same image.

Materials and Methods

Apparatus

The general set-up and apparatus was the same as in Experiment I with the subjects 85 cm from the display and with the 15·0 × 10·0 cm aperture in screen. Four different figures were used as displays. Two were herring-bone figures with 0·6 mm thick lines separated by a perpendicular distance of 3·0 mm and meeting at 55° and 80°, respectively. The third was a figure whose dimensions were those projected by the 55° herring-bone when it was slanted back at 70° to the fronto-parallel plane. These dimensions were ascertained photographically and resulted in a figure whose angles decreased slightly and whose lines got slightly further apart from top to bottom; in the middle of the figure the lines met at 104°. The fourth figure, used as a control, had horizontal lines 0·6 mm thick and separated by 3·0 mm.

Whereas in Experiment I it was assumed that the same figure further away would be perceived as the same size, in this experiment the perceived angle of the herring-bone figures was checked. For this, a white circular board, 20·3 cm in diameter, with two black radial lines was used. One of these lines was made of wire and could be rotated about the end at the centre of the board. Thus the two lines could be set to make any desired angle to each other. The angle could be read off from a scale on the edge of the board, invisible to the subject.

Procedure

Subjects were four male and eight female students. They set the two lines to appear parallel under each of six conditions:

(A) The control figure (horizontal lines) in the fronto-parallel plane.
(B) The same figure slanting back at 70° to the fronto-parallel plane.
(C) The 80° herring-bone in the fronto-parallel plane.
(D) The 55° herring-bone in the fronto-parallel plane.
(E) The 55° herring-bone slanted back at 70° to the fronto-parallel plane.
(F) The projection of condition E in the fronto-parallel plane.

After setting the lines to appear parallel in each of conditions C to F, the subject also set the adjustable angle device so that the angle made by its two radial lines appeared to be the same as that made by the herring-bone lines in the figure. For condition F, where the lines met at different angles at different points on the figure, subject was told to make the setting for the lines in the middle of the figure.

Order effects were controlled in two ways. First, the control conditions A to C were always taken before conditions D to F; each subject had a different combination of the six permutations of order within any group of three conditions. Second, each subject made two settings for each condition going through the six conditions in the order assigned and two more for each coming back through in the reverse order. Movement error for setting

the parallel lines was controlled as in Experiment I. A similar procedure was adopted for setting the angles.

Results

Mean illusion error for each S on each condition was computed as in Experiment I. To allow for any difference in the ability to set lines to appear parallel when they are in the fronto-parallel plane or at a slant, the error for the appropriate control condition A or B was next subtracted from the errors for the other conditions. These were then converted into the angle at which the lines converged. The setting of the radial lines to give the perceived angle of the herring-bone lines was read off direct from the apparatus. The overall means of these for the different conditions are given in Table II.

TABLE II

Results for Experiment II

Condition	Real angle	Projected angle	Mean perceived angle	Mean error: Increased separation of lines at bottom (cm)	Increased separation at bottom corrected (cm)	Angle at which parallels converge
A	180°	180°		0·19		
B	180°	180°		0·11		
C	80°	80°	78·4°	0·70	0·51	1° 4′
D	55°	55°	53·6°	1·60	1·41	2° 52′
E	55°	104°	70·4°	0·08	−0·03	−0° 2′
F	104°	104°	96·4°	0·39	0·20	0° 24′

Statistical analyses were again based on the means for each of the 12 subjects. It can be seen that the apparent angle of the lines in the slanted herring-bone figure (condition E) was greater than in the same figure in the fronto-parallel plane: there was not perfect constancy in the slanted condition. However, it was still seen as having a significantly smaller angle than the 80° herring-bone (condition C) (t for correlated scores = 2·36; $P <0·05$). Looking at the error in setting the vertical lines to appear parallel, it can be seen that it decreases as the angle of the herring-bone increases (conditions D, C and F in that order). The critical condition is the error in the slanted herring-bone: is this more consistent with the perceived angle of the figure (between conditions C and D) or with its projected image (condition F)? In fact, the error in condition E is significantly less than that in condition C ($t = 4·37$, $P <0·01$) and not significantly different from that in condition F ($t = 1·76$, $0·10 < P <0·20$).

Discussion

In both experiments, significant changes in the size of the illusion are produced by changes in the retinal image characteristics in circumstances in which the per-

ceived form of the inducing figure remains relatively constant. In the first experiment, the retinal image is made smaller, first by reducing the size of the figure while maintaining the distance, then by increasing the distance of the constant sized figure. In both cases the illusion is decreased, and by a similar amount, even though the real, and presumably perceived, size of the figure is held constant in the second case. And, in the second experiment, the error in the slanted figure again approximates that in the fronto-parallel projection rather than in its perceived equivalent.

It may be objected that the perceived size of the distant figure has not been checked in the first experiment. The addition of such a check in the second experiment certainly showed less than perfect constancy and, after pilot work had showed this, the 80° condition was added to allow for this. But it seems unlikely that such a control would have invalidated the conclusions drawn from Experiment I. All normal depth cues were available and the display certainly looked very much further away in the distant condition. In fact, for experimental convenience, it was the subject who moved back from the display rather than the display itself which moved away from him: there can be no question that he did not perceive the change in distance. Even if there was not perfect constancy, it is not likely that it fell much short. It is usually much easier to obtain a "real size" match than a "retinal image" match in constancy experiments; and illusion error was very much closer in the distant condition to that in the retinal image equivalent than to that in the "standard" condition.

The results in these experiments are broadly consistent with the results of others in similar situations. Houssiadas and Brown (1963) have found that the error in a number of illusions disappears when they are presented at a sufficient slant or at considerable distances. Marshall and Di Lollo (1963) found a decrease in error in the Hering illusion with an increase in distance; and the same authors, in a later paper (1966), showed the error to be a function of the size of the figure. Gregory (1962) reported an equivalent phenomenon for the Müller-Lyer figure. In this, a figure shaped like a capital "I" was projected onto a vertical rectangular corner and viewed from a point conjugate with the light source. It then projected an I-shaped figure while shape constancy made it appear to be a Müller-Lyer figure, but it showed no distortion of length. The only empirical discard known to this author was provided by Wallace (1966) who found a U-shaped relationship between distance and error in a herring-bone figure: over the distances used in Experiment I, his results showed little or no effect. The fact that his figure was at a different orientation to that used here, with the parallel lines horizontal, may explain the discrepancy.

These results support the prediction from the two channel constancy theory of Gregory (1963). They may also be said to identify the important determinant of changes in error obtained by altering the distance and slant of illusion figures of this type. In this sense, they offer an explanation of these phenomena. The results referred to in the previous paragraph seem to be compatible with the suggestion that what matters is the retinal image projected, irrespective of what the relations in the figure appear to be. For example, the majority of the figures used by Houssiadas and Brown (1963) produced an increase in the projected angle between

the inducing and distorted contours when they were slanted. An increase in this angle generally leads to a decrease in the illusion, as shown over the range 27·5° to 50° in Experiment II and, more systematically, in unpublished experiments by Horrell (1968). The other figures used by Houssiadas and Brown (1963) are more complex in that slant decreases the projected size of some angles in the figure and increases others. It is consistent that these more complex cases tend to be those needing a greater degree of slant to make the illusion disappear: when the angle between the inducing lines and the parallels in the herring-bone figure is about 20° the error is maximal and gets less for smaller angles (Horrell, 1968) and, where slant decreases the projected angle, this only becomes less than 20° at a considerable slant.

One would predict that, when slant reduces the projected angle between inducing and distorted contours within the range 50° to 25°, the illusion should be increased in extent. This can be readily demonstrated by slanting a herring-bone whose lines meet at 80° about an axis along the parallels rather than about the asis at right angles to this. Another clear case is illustrated in the Zöllner illusion (Fig. 3). Rotate it first about axis AB; the projected angles are increased and the illusion

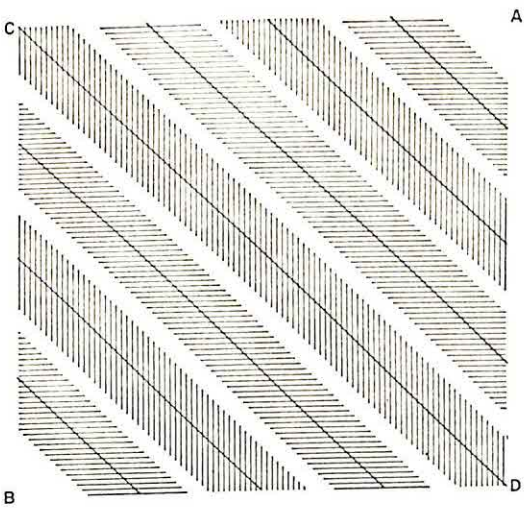

FIGURE 3

reduced. Then slant it about axis CD; the decrease in the projected angle produces an increase in the illusion.

It seems clear, then, that it is the retinal image size or shape which is important rather than perceived characteristics. Just what it is about a smaller size or an increased angle that matters is not indicated by these experiments. The results are certainly not incompatible with such hypotheses as Marshall and Di Lollo (1963), that size has its effect because of an impoverishment of the information extractible from the unit elements of the figure.

Thus, it has been found in these experiments that the error in the herring-bone illusion changes when the distance or the orientation is varied and that these changes are essentially the same as those obtained when the same projected images are reproduced at a fixed distance and orientation. Other workers have found

similar effects of distance and orientation in other illusions. From their results and our knowledge of how configurational variables like angle between lines affect such illusions, it seems reasonable to conclude that the effect of varying distance and orientation is due simply to the changes in the retinal image projected by the figure contours which these manipulations involve. Their results therefore provide evidence to extend the generality of the conclusion that, where the projected form of the contours of a geometric illusion differ from their perceived form, it is the former that determine the size and nature of the illusion. It is argued that Gregory's (1963) theory of geometric illusions would have been incompatible with any evidence to the contrary.

This work was completed while I held a Science Research Council research studentship at Sheffield University. I am very grateful for help given by Professor H. Kay and other members of the Psychology Department at Sheffield and, especially, to Professor R. L. Gregory for his inspiration and encouragement. A preliminary account of some of the results reported here was given at the British Psychological Society conference in London in December 1967.

References

Brown, L. B. and Houssiadas, L. (1965). The perception of illusions as a constancy phenomenon. *British Journal of Psychology*, **56**, 135–42.

Carlson, J. A. (1966). Effect of instructions and perspective drawing ability on perceptual constancies and geometric illusions. *Journal of Experimental Psychology*, **72**, 874–9.

Fisher, G. H. (1968). An experimental and theoretical appraisal of the inappropriate size-depth theories of illusions. *British Journal of Psychology*, **59**, 373–84.

Gregory, R. L. (1962). Visual perception from the engineering point of view—stability and distortions of visual space. Unpublished paper prepared for Internat. Congr. on *Human Factors in Electronics*. Long Beach, California.

Gregory, R. L. (1963). Distortions of visual space as inappropriate constancy scaling. *Nature, London*, **199**, 678–80.

Gregory, R. L. (1968). Perceptual illusions and brain models. *Proceedings of the Royal Society B.*, **171**, 279–96.

Hamilton, V. (1966). Susceptibility to the Müller-Lyer illusion and its relationship to differences in size constancy. *Quarterly Journal of Experimental Psychology*, **18**, 63–72.

Horrell, R. I. (1968). The geometric illusions as inappropriate constancy scaling: an examination of a theory. Thesis submitted for Ph.D. degree at the University of Sheffield.

Houssiadas, L. and Brown, L. B. (1963). The effect of viewing slant and distance on some visual illusions. *Australian Journal of Psychology*, **15**, 108–12.

Marshall, A. J. and Di Lollo, V. (1963). Hering illusion with impoverishment of stimulus in scotopic and photopic vision. *American Journal of Psychology*, **76**, 644–52.

Marshall, A. J. and Di Lollo, V. (1966). Distortion of parallel lines in geometric fields as a function of size of display. *Psychonomic Science.*, **4**, 405–6.

Tausch, R. (1954). Optische Täuschungen als artifizielle Effekte der Gestaltungsprozesse von Grossen- und Formenkonstanz in der naturlichen Raumwahrnehmung. *Psychologische Forschung*, **24**, 299–348.

Wallace, G. K. (1966). The effect of background on the Zöllner illusion. *Acta psychologica*, **25**, 373–80.

Received 6 October 1970

Quarterly Journal of Experimental Psychology (1971) 23, 107–112

ATTENTION IN TRANSCRIPTION SKILL

L. H. SHAFFER

Department of Psychology, University of Exeter

Two viewpoints of attention are compared: one assumes restrictions in a central processor and the other restrictions on simultaneous activity in networks of processes. The distinction between central and distributed processing may not be strongly testable but the distinction between simultaneous and consecutive processing, for different processing functions, is. An analysis of data from a study on learning to type shows very clearly that simultaneous processing of stimulus input and response output is possible.

Homunculus theory lingers in modern theories of attention and performance in the form of a central processor, of limited size, that "makes decisions". In such theories there are usually processes peripheral to the central processor but they are "automatic", do not require attention and so can occur simultaneously. Thus the restrictions of attention occur only at the level of the central processor and the theories diverge on the functions they assign to it (i.e. the functions that are non-automatic). Some assume it is stimulus recognition that requires attention (Broadbent, 1958; Treisman, 1964; Neisser, 1967) and others response organization and transactions with memory (Deutsch and Deutsch, 1963; Norman, 1968).

One could argue that in recent statements of these theories the divergencies are verbal rather than formal (Shaffer and Hardwick, 1969*a*), but more important here is to emphasize that in the absence of operational criteria it is arbitrary to assume that some processes are automatic while others involve decision. It leads to a circular definition of attention, because as soon as processes are shown to co-occur they can be labelled "pre-attentive" or "automatic".

An alternative viewpoint is that processing, between stimulus and response, is distributed in a series (Donders, 1868), or more generally a network (Christie and Luce, 1956) of units. What is called attention may then be thought of as a restriction on the number of (non-convergent) networks or units in a network that can be simultaneously active. In effect the viewpoint makes processing and attention synonymous.

This viewpoint is worth considering only if it can be shown that functionally distinct processes can be simultaneously active. If evidence is to be found it will be in the temporal features of skill and it seems that transcription skills can provide such evidence. Characteristic of these skills is a lag between stimulus and response, observed, for instance, in reading aloud (Buswell, 1927), receiving morse (Bryan and Harter, 1899), typing from script (Butsch, 1932) and shadowing (i.e. repeating aloud) auditory messages (Cherry, 1953). Speed and accuracy of transcription are intimately dependent upon this lag: if in typing the lag is reduced by curtailing stimulus preview then speed falls drastically and errors increase

(Hershman and Hillix, 1965). Also, of course, inter-response times in transcription can be much lower than would be expected from studies of discrete choice-reaction time (Shaffer and Hardwick, 1968). Such observations suggest that processes which operate upon stimuli and organize responses can overlap in time but the gross level of description is consistent also with, for instance, the possibility that the input and output processes are consecutive and that they group stimuli and/or responses.

A critical test for simultaneous input and output processing has been made on data obtained in a recent study of typing. The study is reported more fully elsewhere (Shaffer and Hardwick, 1970) but the analysis here is new. The test is based upon the following assumptions.

Let us assume two stages of processing, one of stimulus recognition the other of response organization: call them input and output stages and let them have random latencies τ and t, respectively. The total time to transmit a single letter of text is $T = \tau + t$ and this is the time per response that is observed if text is exposed one letter at a time.

Grouping letters at input is effective in increasing speed if, for groups of size k, $\tau_1 \leqslant \tau_k < k\,\tau_1$; similarly grouping responses at output is effective if $t_1 \leqslant t_k$, $t_k + (k - 1)t < kt_1$. (A variable, τ or t, is given a subscript if it initiates a group.) It is assumed that if response groups occur they preserve stimulus groups.

If both stages can operate cyclically on their respective inputs there must be a buffer store between them to hold intermediate codes, which are the output of the first stage and input to the second. In this case a situation can develop in which a queue of codes is continually renewed from the input stage and served by the output stage. The latter stage produces typing responses at a rate $T = (\tau_k - \lambda) + t$, where λ is the temporal overlap between initiating a new stimulus input and serving an existing queue, and the bracketed term reduces to zero as long as a queue is maintained.

The hypothesis to be tested, then, is that λ is always zero. This means that no new input is initiated until the output process has emptied the buffer store and implies that one should find response latencies at least of magnitude $T = \tau_1 + t_1$ alternating with runs of $k - 1$ response latencies $T = t$.

Method

Thirty subjects learned to type on a special keyboard. It had the letters ACDEHILNORST, arranged in two rows, together with a space bar. This keyboard had automatic control of display renewal and data recording via transistorized circuits. The display was projected by a bank of six in-line units and a shift register transposed the display symbols from right to left, by one symbol, immediately a key was pressed. The number of symbols actually on display in any trial could be reduced by use of shutters. Since the subject had always to type the leftmost symbol the others on display were available for preview. The apparatus provided a complete record of all response latencies and errors.

Three kinds of text were constructed resembling English at the level of a word, syllable or letter: W text contained only dictionary words put into semi-prose sequences; S text contained only words having a strict vowel-consonant alternation, but not forming dictionary words; L text contained only words that were either all vowels or all consonants. Each text used word lengths of 1 to 6 letters seven times, giving 42 words in a text, or 188

symbols including spaces. A large number of different texts were constructed from a vocabulary of over 1000 words and a subject would not encounter the same text more than two or three times in the whole experiment.

The data to be presented was obtained in the thirteenth session of practice, each of the previous sessions lasting about 1 hr in which usually 16 texts were typed. In this session different texts of each kind were presented with levels of exposure 1, 2, 4 or 6 symbols. The first of these is called the no preview (NP) condition and the others preview (P) conditions.

Results

Frequency distributions of response latencies were obtained, for each subject, for each kind of text and amount of preview. They were typically as in Figure 1,

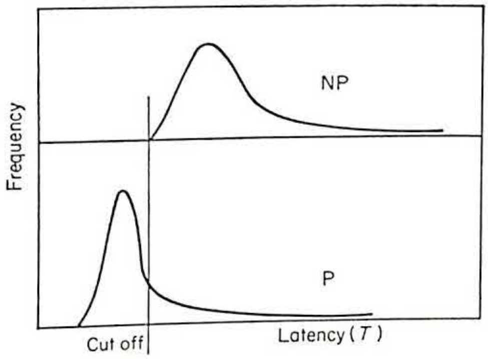

FIGURE 1

for conditions of no preview (NP) and preview (P). From the above considerations the distribution NP is that of $T = \tau + t$, while P is a mixed distribution of $T = t$ and $T = (\tau - \lambda) + t$. (We ignore the suffixes that indicate group size and in doing so bias the following test in favour of the hypothesis $\lambda = 0$.)

The test proposed is to look for the occurrence of runs of fast responses of length $r \geqslant k$, where k is the size of text exposure in that trial. This is a critical test since at least one new grouping must be initiated on any k consecutive stimuli: all that is required is a cutoff point in the distribution of P to define a fast response and it is sufficiently stringent to take the fastest latency achieved in NP. This test was performed on each subject's data.

First we show the locations of the peaks in P and NP in relation to the cutoff point, so as to indicate the safety margin of the test. In Table I, averaged over all

TABLE I

Reference latencies in the NP and P distributions

Text	L1	L2	L3	L4	L5
W	350	670	280	250	250
S			290	270	270
L			300	280	280 msec

subjects, L1 is the minimum latency in NP and L2 the median latency in NP. These are composite figures for all kinds of text since the effect of text without preview is negligible. L3, L4 and L5 are the medians of all latencies below cutoff in P_k, $k = 2$, 4 and 6, respectively. In all cases the peak in P lay well below the cutoff and the peak in NP well above it.

TABLE II

Frequencies of runs of fast responses

Text	P_k	Run length, r 1	2	3	4	5	6	7	8	9	10	11	12	>12
W	6	560	284	233	100	65	29	26	15	10	12	3	5	17
	4	607	308	169	100	51	32	23	13	10	7	1	2	5
	2	977	251	52	21	8	3	2	1	0	1	0	0	1
S	6	826	350	99	36	30	11	8	1	2	2	2	1	3
	4	768	289	109	63	22	11	5	6	5	3	0	1	3
	2	725	113	26	14	7	9	2	2	0	1	0	0	1
L	6	766	284	114	34	22	17	5	4	5	1	5	0	3
	4	733	269	108	56	28	16	10	4	6	2	1	7	4
	2	717	132	23	18	7	4	1	1	1	0	2	1	0

Next we show in Table II the frequencies of runs of length r of fast responses in P_k. It is seen that there were frequent runs of size $r \geqslant k$, the longest observed run being 26. To give these figures a perspective the percentage of all response latencies that appear in fast runs of length $r \geqslant k$ is shown in Table III. On the hypothesis that $\lambda = 0$ these percentages should be zero.

TABLE III

Percentage response latencies occurring in runs of length r $\geqslant P_k$

	P_k 6	4	2
Text W	19	25	15
S	5	11	9
L	6	14	9

The error rate ranged between 2 and 4%, the higher rates appearing on nonsense texts or with reduced preview. Some of these errors occurred in the runs of fast responses recorded in Table II: removing such runs would require minor changes in the Table and would not materially affect the result.

Discussion

We have provided a test of whether two functionally distinct processes can be simultaneous, not of whether such processes are structurally distributed or handled by the same processor. It is unlikely that the latter distinction can be critically

tested by behavioural data: one can only argue that it is conceptually simpler to suppose that simultaneous processes occur in structurally distinct units.

The data were obtained at a stage of moderate keyboard practice, while the subjects were still very much slower than skilled typists (cf. Shaffer and Hardwick, 1968), nevertheless they are sufficient to reject the hypothesis that $\lambda = 0$. In terms of the transcription model outlined above one may suppose that a skilled typist can sustain a queue of codes in the buffer store and so produce a monotonous stream of fast responses, and the cumulative records obtained from typists by Genest (1956) are consistent with this.

A point to be dealt with here is the misconception that in transcription people not only read ahead in the text (musical score, etc.) but they also anticipate what is to come next. Although subjects can predict the sequel to an incomplete sentence (Garner, 1962; Morton, 1964) there is no evidence that they utilize such prediction in transcription. On the contrary skilled typists type prose no faster than random word sequences (Shaffer and Hardwick, 1968). In a later part of the experiment discussed here it was shown (Shaffer and Hardwick, 1970) that giving postview of six letters of text led to no faster performance than presenting text one letter at a time, although postview displayed as much information, as a basis for prediction, as preview. Thus one may not suppose that in the present data long runs of fast responses were achieved by predicting ahead of the display. In any case the S and L texts were not predictable beyond the general vowel-consonant structure within a word.

As it has been described the transcription model gives only a minimum account of the processing involved and does not specify details of the processing stages (cf. Quastler and Wulff, 1955; Shaffer and Hardwick, 1970). It does not include the monitoring processes that enable the detection of errors as they occur (Shaffer and Hardwick, 1969*b*); or consider features of expressive modulation in reading aloud or playing music. Thus there is probably far more simultaneous processing than has been revealed in the present result.

It has been claimed that experiments on "refractoriness" have demonstrated non-overlap in the processing of successive stimuli (Welford, 1967; Davis, 1957). In fact few of the experiments have been properly designed to test this assertion (Shaffer, 1968), but it is true that there is sometimes delay of response to the second of two stimuli. The magnitude of delay may be consistent with an assumption that the input and output units operate sequentially upon their respective inputs, but is usually too small to satisfy an assumption that processing in these units do not overlap in time. The theories of Welford and Davis seem to make both assumptions and are contradicted both by the present results and those obtained in "refractoriness" experiments.

We advocate that the study of attention should take account of the overall configuration of processing in skills. It enables a more comprehensive approach if response commitment as well as stimulus parameters are varied. For example, much of the present work on transmitting information in multiple inputs has used the shadowing method with dichotic messages (Treisman, 1964) and little emphasis has been given to the fact that the subject is required to listen and speak at the same time. A different estimate of how much of the messages he perceives

can be obtained if a probe method, entailing only occasional motor response, is used (Shaffer and Hardwick, 1969*a*).

I would like to thank Miss Jane Hardwick and Mrs Ann French for their assistance and the S.R.C. for financial support.

References

BROADBENT, D. E. (1958). *Perception and Communication.* Oxford: Pergamon.

BRYAN, W. L. and HARTER, N. (1899). Studies on the telegraphic language. *Psychological Review*, **6**, 345–75.

BUSWELL, G. T. (1927). An experimental study of the eye-voice span in reading. *Supplement Education Monograph* 17.

BUTSCH, R. L. C. (1932). Eye movements and the eye-hand span in typewriting. *Journal of Educational Psychology*, **23**, 104–21.

CHERRY, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, **25**, 975–9.

CHRISTIE, L. S. and LUCE, R. D. (1956). Decision structure and the time relations in simple choice behaviour. *Bulletin of Mathematical Biophysics*, **18**, 89–111.

DAVIS, R. (1957). The human operator as a single channel information system. *Quarterly Journal of Experimental Psychology*, **9**, 119–29.

DEUTSCH, J. A. and DEUTSCH, D. (1963). Attention: some theoretical considerations. *Psychological Review*, **70**, 80–90.

DONDERS, F. C. (1868). Die Schnelligkeit psychischer Processe. *Archiv für Anatomie und Physiologie*, 657–81 (translated in *Acta Psychologica* (1969), **30**, 412–31).

GARNER, W. R. (1962). *Uncertainty and Structure as Psychological Concepts.* New York: Wiley.

GENEST, M. (1956). D'analyse temporelle du travail dactylographique. *Bulletin du Centre d'études et de recherches Psychotechnique*, **5**, 183–91.

HERSHMAN, R. L. and HILLIX, W. A. (1965). Data processing in typing. *Human Factors*, **7**, 483–92.

MORTON, J. (1964). The effects of context on the visual duration threshold for words. *British Journal of Psychology*, **55**, 165–80.

NEISSER, U. (1967). *Cognitive Psychology.* New York: Appleton-Century-Crofts.

NORMAN, D. (1968). Towards a theory of memory and attention. *Psychological Review*, **75**, 522–36.

QUASTLER, H. and WULFF, V. J. (1955). Human performance in information transmission. Report R-62, Univ. of Illinois.

SHAFFER, L. H. (1968). Refractoriness in information processing. *Quarterly Journal of Experimental Psychology*, **20**, 38–50.

SHAFFER, L. H. and HARDWICK, J. (1968). Typing performance as a function of text. *Quarterly Journal of Experimental Psychology*, **20**, 360–9.

SHAFFER, L. H. and HARDWICK, J. (1969*a*). Monitoring simultaneous auditory messages. *Perception and Psychophysics*, **6**, 401–4.

SHAFFER, L. H. and HARDWICK, J. (1969*b*). Errors and error detection in typing. *Quarterly Journal of Experimental Psychology*, **21**, 209–13.

SHAFFER, L. H. and HARDWICK, J. (1970). The basis of transcription skill. *Journal of Experimental Psychology*, **84**, 424–40.

TREISMAN, A. M. (1964). Selective attention in man. *British Medical Bulletin*, **20**, 12–16.

WELFORD, A. T. (1967). Single-channel operation in the brain. *Acta Psychologica*, **27**, 5–22.

Received 25 June 1970

Quarterly Journal of Experimental Psychology (1971) 23, 113–117

ADAPTATION TO DISPLACED VISION: REAFFERENCE IS A SPECIAL CASE OF THE CUE-DISCREPANCY HYPOTHESIS

BERNARD MOULDEN†

Department of Psychology, University of Bristol

Subjects were exposed to a prismatically displaced view of their actively-moved right hand which was optically "stopped"; they achieved as much adaptation in this condition as in one in which they were allowed full "reafferent" stimulation. This provides further evidence against Held's "reafference" hypothesis.

Introduction

Much recent research has been devoted to the problem of the necessary antecedent conditions for adaptation to prismatically-displaced vision. The "reafference" hypothesis, introduced into this field by Held (Held and Hein, 1958; Hein and Held, 1962; Held and Freedman, 1963) was first formulated by Holst and Mittelstaedt (1950) who rearranged the visual input of insects and fish and concluded that the necessary condition for adequate visual-motor co-ordination is the perception of the relation between actively-produced movement of body parts and the changes in sense-organ stimulation consequent upon such movement; such changes in sensory stimulation contingent upon self-produced movement they called "reafference", while stimulation of the sense organs produced solely by non-contingent changes in the external world was called "exafference".

Stratton (1897), Kohler (1962), Wooster (1923), and other early workers have all emphasized the importance of active movement to the process of adaptation to optical rearrangement, but Held was the first to formalize the notion in hypothesizing that the necessary antecedent conditions for visual-motor co-ordination or re-co-ordination were the continuous comparison of visual and proprioceptive information resulting from *self-produced movement*. Held and his co-workers have presented considerable experimental evidence which is claimed to support this hypothesis. The experiment of Held and Freedman (1963) is typical: the subject viewed his hand through a laterally-displacing prism for 3 min while the hand was (a) motionless, or (b) moved passively from side to side by the experimenter, or (c) moved actively. Only this last condition produced any adaptation as measured by comparing accuracy on a pointing task before and after exposure.

Wallach and Karsh (1963), and Howard, Craske and Templeton (1965), suggested an alternative to the reafference hypothesis; Wallach and his colleagues call it a "cue-discrepancy hypothesis"; Howard *et al.* talk of "discordant exafferent stimulation" which will produce changes in the localizing behaviour of a passive subject if it is sufficiently "strong". For the purposes of this argument it is not

† Present address: Department of Psychology, University of Reading.

necessary to concern ourselves with the difficulty of specifying the "strength" of discordant cues except post hoc; the central point is that these two groups of workers both suggest that the sufficient antecedent condition for the occurrence of adaptation is "salient" or "powerful" information—*regardless of its source*—about the non-correspondence between the visually perceived locus of objects and where one perceives them to be via some modality other than vision—typically, the haptic system.

It is clear that the two hypotheses are not mutually exclusive ("reafference" could be subsumed under "cue-discrepancy"); but the "reafference" hypothesis, if correct, would be more specific, and hence more powerful, and the purpose of this paper is to provide a further test for the more specific "reafference" hypothesis.

The two hypotheses are made even less distinct by an experiment reported by Kravitz and Wallach (1966). Held had reported (1958) that merely viewing one's stationary arm through a prism did *not* produce adaptation; Kravitz and Wallach, however, reported that adaptation *did* occur if the passive arm was subjected to low amplitude, high-frequency vibration during viewing. They suggested that "both self-produced movement and vibrating stimulation operate by enhancing the central effect of proprioceptive stimulation representing the true arm position". If this suggestion is correct, then it is quite possible that the importance of self-produced movement lies not in the fact of movement—continuous translation through space—itself, but in some concomitant information-enhancing function. This experimental hypothesis may be tested by removing from the situation all visual information about the *movement* of the limb and providing, as the only visual information contingent upon self-produced movement, data concerning only the *instantaneous arm position*. If adaptation occurs under these circumstances the "one-to-one relation between movement and its sensory feedback" (Held, 1963) will have been shown not to be necessary.

Method

Apparatus

The apparatus consisted of a black-painted table through the top of which protruded a vertical lever, pivoted about a point 16 in. below the surface of the table and free to move against a light spring loading in a slot cut in the table (see Fig. 1). Subjects were instructed to grasp the lever and move it, to the beat of a metronome, once every 2 sec, from side to side. Subjects, who were all right-handed, used their right arm which was thus constrained to move in a 45° arc of radius 16 in. in a plane normal to the line of sight.

To the bottom of this lever was attached a microswitch which fired a dim stroboscope in the "external trigger" mode, so that a light flash, specified by the manufacturers (Dawe Instruments Ltd.) to be between 4–12 sec$^{10-6}$, occurred only when the lever was in the vertical position; consequently, subjects saw their hand, which was in fact moving from side to side at the rate of once every 2 sec, always apparently in the same place.

It will be seen that this design satisfied the condition that subjects should be making self-produced movements while receiving visual information only about one instantaneous position of their hand.

The experiment was carried out in a sound-proof, light-free room, the only extraneous stimulus being the sound of a ventilation fan. Before and after each of the four exposure conditions, each subject was given tests consisting of five attempted localizations of each of three targets presented in a test-box of the kind designed by Held and Gottlieb (1958).

This consisted of a box in which subjects saw the virtual image of targets which appeared to lie in a horizontal plane below a mirror. They were required to mark the apparent location of the virtual images, using a different-coloured felt-tipped pen for pre- and post-tests. This testing procedure was adopted because the mirror obscured both the subject's hand and his markings and thus prevented him from recognizing and correcting his errors.

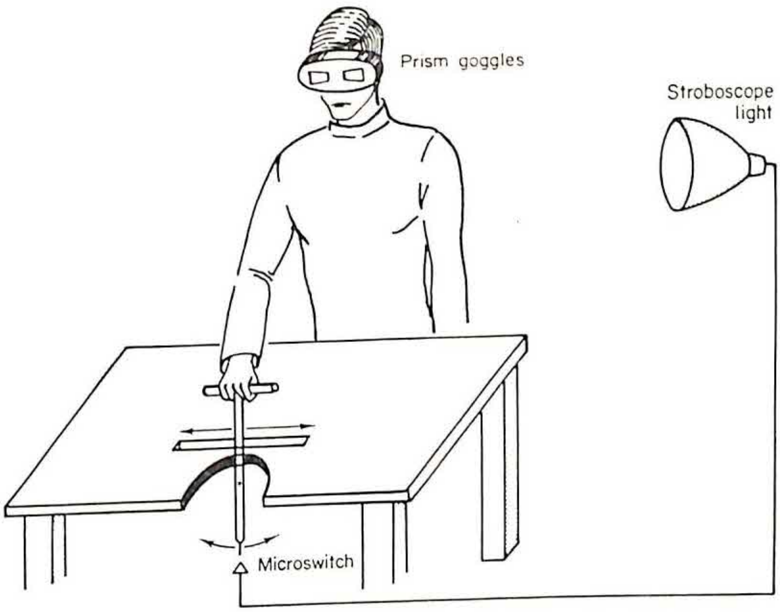

FIGURE 1

No subject was permitted to look through the prisms under conditions of full illumination except in condition 4, which was the last condition for all subjects (see under *Experimental Design*).

For each subject the mean error of localization on the horizontal axis was measured in inches for pre- and post-test; the pre-test score for each target was subtracted from the corresponding post-test score; the score for each of the three targets was averaged and this "average after-effect" used as an index of adaptation.

Subjects

Twelve subjects, eight female and four male, all undergraduate psychology students who were naïve regarding the purpose of the experiment, were employed.

Experimental design

Each subject experienced each condition and thus acted as his own control. A 10 min rest period separated each condition from the previous one, and during this period the subject was encouraged to move around the well-illuminated laboratory and to pick up and manipulate objects to ensure re-adaptation to normal before experiencing the next condition.

There were four conditions:

(i) *Experimental* 1 (E_1). Subject wore 20° laterally displacing prisms mounted in modified underwater goggles; exposure was to 3 min of arm rocking with contingent flash illumination.

(ii) *Control* 1 (C_1). As above, but with no illumination (to control for fatigue effects from the arm movement).

(iii) *Control* 2 (C_2). Wearing prisms, subjects viewed their hand resting upon the lever which was rigidly locked in the upright position, while the stroboscope flashed once per 2 sec for 3 min.

(iv) *Experimental* 2 (E_2). As condition (i), but with full room lighting and no flash; this was the "normal reafference" condition, and is very similar to one used by Held and Gottlieb (1958).

All exposure periods were split into two 90 sec sessions by a 30 sec rest period in darkness. (Subjects in a pilot experiment had complained that their arm ached as a result of 3 min of continuous movement.)

Half of the subjects wore base-right, and half base-left prisms for all conditions. Each subject experienced condition (iv) last. This procedure was adopted—since it is known that adaptation *would* occur in this situation—to rule out the possibility of any "contamination" of subsequent experimental or control conditions by previous adaptation. The three remaining conditions could be presented in six different orders; each order was assigned to one base-right subject and one base-left subject, so that each of 12 subjects received one of the 12 possible combinations of order and direction of displacement.

Results

Results are summarized in Table I; the results for subjects with base-right and base-left prisms are pooled and the after-effect measures are given in degrees. The data were analysed by a subjects × treatments analysis of variance. The only significant difference was that between the experimental conditions and the control conditions ($F = 23·63$, $df = 1, 23$, $P < 0·01$); the experimental conditions E_1 and E_2, were not significantly different ($F = 0·19$, df 1, 11, $P > 0·05$); nor were the control conditions C_1 and C_2 ($F = 0·36$, df 1, 11, $P > 0·05$).

TABLE I

Means and standard deviations of after-effects in degrees

Condition	E_1 "Stopped" viewing	E_2 Full viewing	C_1 Arm moving, no viewing	C_2 Arm still, stroboscopic viewing
Mean	+7·13	+7·46	+1·48	−1·15
S.D.	3·55	2·71	2·92	3·72

A positive sign represents an after-effect in the direction opposite to prism deviation.

It can be seen from the table that the two control conditions produce no significant adaptation, while the two experimental conditions, active arm movement in full room lighting or in single-shot strobe exposure, produce degrees of adaptation which are not significantly different from each other.

Discussion

After-effects of adaptation to laterally-displaced vision were obtained in conditions in which the subject did not experience "the one-to-one relation between movement and its sensory feedback" (Held and Freedman, 1963), which Held further characterizes thus: ". . . each distinguishably different movement of the hand will be accompanied by a unique change in the viewer's image of that hand" (Held and Freedman, 1963), and which his reafference hypothesis implies to be a necessary precondition for adaptation. We may thus conclude that the present experiment provides a further disproof of this hypothesis. The experiment

reported here supports the suggestion that the importance of self-produced movement lies not in the fact of movement but in the fact of efference; as a general conclusion one might say that Held's admirably seminal "reafference" principle should be replaced by an "efference-enhanced afference" principle (which is simply reafference without the movement-monitoring principle) and that this may be subsumed under the cue-discrepancy hypothesis.

Bearing in mind the results of this experiment and that of Wallach and Kravitz, it is interesting to speculate upon the neurophysiological events underlying the observed behavioural ones; it would be very significant, for example, if both vibration and active limb movement were found to produce increased neural firing in that part of the post-central gyrus which receives innervation from the joint capsules and pericapsular tissue in monkeys (Mountcastle and Powell, 1959).

The above experiment was carried out at the Department of Psychology, University of Bristol, while I was in receipt of an S.R.C. studentship. I wish to thank Dr Stuart Anstis, of Bristol, and Dr Brian Craske, of Southampton, for valuable discussion, advice, and encouragement; and Mr Hadyn Ellis of Reading for his criticism of the manuscript.

References

HARRIS, C. S. (1965). Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review*, **72**, 419–444.

HEIN, A. V. and HELD, R. M. (1962). A neural model for labile sensorimotor co-ordinations. In *Biological Prototypes and Synthetic Systems*, Vol. 1. New York: Plenum Press.

HELD, R. M. and FREEDMAN, S. J. (1963). Plasticity in human sensorimotor control. *Science, N.Y*, **142**, 455–62.

HELD, R. M. and GOTTLIEB, N. (1958). Technique for studying adaptation to disarranged hand-eye co-ordination. *Perceptual and Motor Skills*, **8**, 83–6.

HELD, R. M. and HEIN, A. V. (1958). Adaptation of disarranged hand-eye co-ordination contingent upon reafferent stimulation. *Perceptual and Motor Skills*, **8**, 87–90.

VON HOLST, E. and MITTELSTAEDT, M. (1950). Das Reafferenzprinzip. *Naturwissenschaften*, **37**, 464–76.

HOWARD, I. P., CRASKE, B. and TEMPLETON, W. B. (1965). Visuomotor adaptation to discordant exafferent stimulation. *Journal of Experimental Psychology*, **70**, 189–91.

KOHLER, I. (1962). Experiments with goggles. *Scientific American*, **206**, 62–86.

KRAVITZ, J. H. and WALLACH, H. (1966). Adaptation to displaced vision contingent upon vibrating stimulation. *Psychonomic Science*, **6**, 465–6.

STRATTON, G. M. (1897). Vision without inversion of the retinal image. *Psychological Review*, **4**, 341–60, 363–481.

WALLACH, H. and KARSH, E. B. (1963). The modification of stereoscopic depth-perception and the kinetic depth-effect. *American Journal of Psychology*, **76**, 429–35.

WOOSTER, M. (1923). Certain factors in the development of a new spatial co-ordination. *Psychological Monographs*, **32**, No. 4 (whole No. 146).

Received 21 *May* 1970

Quarterly Journal of Experimental Psychology (1971) 23, 118–125

AN ANALYSIS OF OVERSHADOWING AND BLOCKING

N. J. MACKINTOSH
Dalhousie University, Halifax, Nova Scotia, Canada

In classical aversive conditioning experiments, rats do not always learn about all aspects of a compound stimulus predicting shock. A strong stimulus may overshadow a weaker one; and pretraining on one component may block learning about a second component. These results have been explained either by appealing to a notion of selective attention, or by assuming that learning about one component is a function of prior response strength to the entire compound of which it forms a part. In Experiment I, overshadowing was demonstrated on the first trial of conditioning, i.e. before either component had acquired any response strength. In Experiment II, pretraining on one component resulted in complete failure to learn about a second component during compound training, but did not prevent additional learning about the first component. Both results were interpreted as supporting an attentional analysis of blocking and overshadowing.

Introduction

The term "overshadowing" was used by Pavlov (1927) to describe the following phenomenon: if an animal was conditioned to a compound CS consisting of one relatively strong and a second relatively weak component, test trials with the two components in isolation revealed conditioning to the stronger element but none to the weaker. Although the weak element served as an adequate CS when an animal was trained with it in isolation, in a compound with a stronger element it was completely overshadowed. There is now evidence that this sort of result, although not necessarily obtained in such an extreme form, occurs in a wide variety of situations—including classical aversive conditioning (Kamin, 1969), operant discrimination learning (Miles, 1969) and simultaneous discrimination learning (Lovejoy and Russell, 1967)—and that the magnitude of the overshadowing effect depends upon a number of variables including the relative salience of the two components (Miles, 1965), and their reinforcement schedules (Wagner, Logan, Haberlandt and Price, 1968). One variable whose effects have been dignified by a special title is the extent of prior training with one of the elements in isolation. Again in a variety of situations—including classical conditioning (Kamin, 1969), operant (Johnson, 1970; Mackintosh and Honig, 1970) and simultaneous (Mackintosh, 1965) discrimination learning—it has been found that pretraining on one component of a subsequently presented compound cue increases the extent to which that component overshadows the other. This effect has been called "blocking" by Kamin, but as most writers on the subject have agreed, it is likely to be closely related to the other phenomena subsumed under the term overshadowing.

The most obvious interpretation of these data is in terms of theories of learning that incorporate the concept of selective attention (Sutherland and Mackintosh, 1971). According to such theories, stimuli compete for the subject's limited attention, the greater the extent to which the subject attends to one cue the less likely he is to attend to another, and overshadowing is a simple consequence of the presence of a second cue reducing the extent to which the subject can attend to and hence learn about the first. Variables that may be assumed to increase the extent to which the subject attends to the overshadowing cue (such as its salience, relevance and the subject's prior experience with it) will of course increase the extent to which it overshadows another cue.

A second general explanation of overshadowing phenomena was recently suggested by Kamin (1969) and has been more formally advanced by Rescorla (1969), Wagner (1969) and Rescorla and Wagner (1971). The idea behind this explanation can be most easily understood by considering a classical conditioning experiment on blocking. If a rat experiences a number of pairings of a stimulus, A, and electric shock, and then a number of compound conditioning trials in which the shock is signalled by A plus another stimulus, B, pretraining on A will often completely block conditioning to B. Kamin suggested that no conditioning occurs to B because on trials when B occurs (in conjunction with A) the occurrence of shock is already fully predictable on the basis of A alone. Only surprising reinforcers are reinforcing. Rescorla and Wagner have put forward essentially the same account: no conditioning occurs to B on AB trials because conditioning is already asymptotic to A. In general, increments in response strength to any stimulus are a decreasing function of current response strength to the entire compound of which the stimulus forms a part. If the strength of the AB compound is high, little conditioning can occur to B. The argument can be generally applied to all overshadowing data, since conditioning will occur rapidly to the more salient component of a compound CS, response strength will rapidly approach asymptote, and there will be none available for conditioning to the weaker component.

On this analysis, overshadowing is assimilated to blocking rather than vice-versa: blocking occurs because response-strength is already asymptotic before the second component is added; overshadowing occurs because response-strength approaches asymptote more rapidly with a compound CS than with a single CS. One implication of this analysis is that overshadowing should not occur on the very first trial of training: if overshadowing of stimulus B by stimulus A is due to the presence of A increasing response-strength of the AB compound and thereby reducing the magnitude of further possible increments to B, no overshadowing can occur when both components have zero response strength, i.e. on the first training trial. An attentional analysis, on the other hand, although it might predict less overshadowing on Trial 1 than on later trials, basically sees overshadowing as due to competition between stimuli, and will certainly predict some on Trial 1.

Experiment I

To discover whether overshadowing will occur on the first trial of training, it is necessary to have a situation in which substantial learning occurs in a single

trial. If a stimulus is paired with shock on a single trial, then subsequent presentation of the stimulus suppresses licking at a water tube in thirsty rats (e.g. Vogel and Spear, 1966), and this was the procedure used in the present experiment. A second problem was to control for the effects of generalization decrement. Overshadowing is assumed to have been demonstrated if a group trained with AB and tested with B alone shows less suppression to B than does a group trained with B alone. But for the first group, there is a change from training to testing (namely, removal of A), and this alone might be sufficient to disrupt test performance (Sutherland and Andelman, 1967). The control adopted here was to train two pairs of groups, one member of each pair with AB, the other with B alone. For one pair of groups, B was a relatively weak tone, which, it was expected, would be overshadowed by A (a relatively strong light); for the other pair, B was a stronger tone. There is good evidence that strong stimuli are not easily overshadowed (Pavlov, 1927; Miles, 1965; compare Sutherland and Andelman, 1967, with Lovejoy and Russell, 1967). If the strong tone is less subject to overshadowing by the light than the weak tone, then we should expect less difference between this pair of groups than between the first. Since, however, in each case, the AB groups have been subject to the same amount of generalization decrement on the test (produced by removal of A), such an interaction could only be attributed to an overshadowing effect.

Method

Subjects and apparatus

The subjects were 40 male hooded rats, obtained from Les Elevages Medico, Montreal, between 250 and 300 g. Ten days before the start of the experiment, they were placed on a gradually stricter schedule of water deprivation: throughout the experiment, they received 1 hr free access to water 30 min after each day's training.

The apparatus was a 9 × 6 × 9 in. high box placed in a sound-insulated, ventilated chamber. Three sides of the box were made of aluminium sheeting, the fourth side was a clear Perspex door. The floor was made of 17 stainless steel rods spaced 1 in. apart, and the ceiling of white Perspex. From the centre of one end wall, a water nozzle, connected to a solenoid operated valve, protruded ½ in. into the box, 2 in. above the floor. General illumination of the box was provided by a 110 V 10 W light, switched through a 2000 ohm resistor, and positioned above the Perspex roof. The stimuli used as CSs, referred to as light and tone, were as follows: the light was an increase in illumination caused by switching on the overhead light at its full intensity; the tone was either a 60 or 83 dB 2000 Hz click produced by a Massey-Dickinson click generator from a loud speaker mounted in the end wall above the drinking nozzle. Licking responses were recorded by a drinkometer circuit by the contact of the rat's tongue with the end of the nozzle.

Procedure

On each of days 1 and 2, subjects received 25 min sessions in the box, with water available on a random interval 18-sec schedule: on average every 18 sec a reinforcement was set up, and the next lick on the nozzle delivered approximately 0·1 cc of water. On day 3, pretest trials were given: during the course of the session, subjects were exposed to one 30 sec presentation of the light, one 30-sec presentation of the tone (either 60 or 83 DB depending on the group), and one or two presentations of the tone-light compound. Licking responses were recorded during the 30 sec preceding stimulus onset (b), and during the 30 sec of stimulus presentation (a), and a suppression ratio $a/a + b$ was calculated for each subject. If the suppression ratio was less than 0·30 on the first compound

trial, a second compound was presented. Only nine subjects required a second trial. Stimulus presentations were programmed to fall at least 5 min apart.

On day 4, the drinking nozzle was removed, and each subject was placed in the box for 5 min. After 2 min, a 30-sec stimulus was presented, immediately followed by a 1-sec 1 MA shock delivered by a scrambler to the grid floor and walls of the box. On day 5, subjects received a 25-min retraining session with the drinking nozzle in place and water again available. On day 6, all subjects received two test trials with the tone, and on day 7, two test trials with the light. No shocks were given on test trials. Stimuli were presented for 30 sec at 10 min apart, and licking responses were recorded during stimulus presentations and for 30 sec preceding each stimulus.

There were four groups in the experiment, distinguished by the stimulus paired with shock: group T was trained with the 83 dB tone, group t with the 60 dB tone, and groups TL and tL with the tone-light compounds.

Results

The results are shown in Table I. Suppression ratios were calculated for each subject's last compound pretest trial, and for the two tone test trials, and two light test trials combined. Pretest suppression ratios were relatively similar for the four groups, and did not depart significantly from 0·50—indicating equal response rates before and during stimulus presentation. On test trials for tone, group tL showed relatively little suppression compared to group t, whereas groups TL and T showed identical suppression ratios. The difference between groups tL and t was significant at the 0·05 level ($t(18) = 2{\cdot}29$). On the light test trials, groups tL and TL were, not surprisingly, more suppressed than the two groups conditioned to the tone alone: each compound group differed from its control ($t(18) = 3{\cdot}00$ and $3{\cdot}41$, $P < 0{\cdot}01$, for groups tL and TL respectively) but the two compound groups did not differ from each other ($t(18) = 1{\cdot}77$, $P > 0{\cdot}05$).

TABLE I

Experiment I: Mean suppression ratios on pretest and test trials

Group	Pretest Compound	Test Tone	Test Light
tL	0·47	0·40	0·34
t	0·47	0·22	0·52
TL	0·51	0·19	0·22
T	0·49	0·19	0·49

The results showed that a weak auditory stimulus can be overshadowed by a stronger visual stimulus on the first (and only) conditioning trial. The similarity of suppression to the tone by groups TL and T suggests that generalization decrement alone (removal of the light) was not the cause of the absence of suppression in group tL, and also confirms that it is only relatively weak stimuli that are easily overshadowed. It should be noted that there was no suggestion that even the strong tone would overshadow the light: group TL was rather more suppressed to the light than was group tL.

Experiment II

Experiment I was designed to test Rescorla and Wagner's analysis of overshadowing. Experiment II was designed to test their analysis of blocking. According to their account, blocking of one component, B, of a compound stimulus, AB, occurs when sufficient training on A alone has been given to bring its response strength close to asymptote. Essentially complete blocking (observed by Kamin, 1969) will occur if response strength is at asymptote before B is introduced. According to an attentional analysis, total blocking simply depends upon *attention* to A being sufficiently strong to preclude attention to, and hence learning about, B. This does not imply that no further learning could occur during compound training, only that any further increases in response strength will accrue to A rather than to B. Rescorla and Wagner assume that if total blocking occurs this can only be because no further increments in response strength, to A or B, are possible. Experiment II tested this prediction. Using the same general procedure as in Experiment I, one group of rats was given initial training on A followed by extended compound training on AB, and was tested for suppression to A and B alone. A second group received the same initial training on A, followed by a single AB compound trial, and the same sequence of test trials. Comparison of the two groups' suppression ratios to B will reveal whether anything was learned about B during extended compound training (i.e. whether total blocking occurred), while comparison of suppression ratios to A will reveal whether extra training increased response strength to A on compound trials. A third group received the same number of compound trials as the first, but without prior training on A. This was to ensure that B was an effective stimulus.

Method

Subjects and apparatus

The subjects were 30 male rats of the same weight and stock, and maintained on the same schedule of water deprivation, as those used in Experiment I. The apparatus was the same as in Experiment I, although the stimuli and shock used differed slightly. The two stimuli were a 70 dB tone, and a light produced by switching on the overhead light at full intensity. The background intensity, however, was brighter than in Experiment I, being produced by switching the light through a 1000 ohm resistor. All stimulus presentations were of 30-sec duration. The shock was a 0·5 sec 1 MA scrambled grid shock.

Subjects initially received 2 days pretraining on the licking response, and on day 3 were given four pretest trials, two to the light, two to the tone. Stimulus presentations were programmed to occur at not less than 5 and not more than 8 min intervals. On day 4, the water nozzle was removed, and the subjects were divided into three groups to receive conditioning trials. With one exception, four CS-UCS pairings were given in each 25-min session with between 5 and 8 min between presentations. Group T + 16 TL received 8 tone trials over days 4 and 5, and 16 tone plus light trials over days 6 to 9. Group T + 1 TL received 8 tone trials over days 4 and 5 and on day 6 a single tone plus light trial (their training session on this day was reduced to 10 min). Group 16 TL received 16 compound tone plus Light trials over days 4 to 7.

After each group had completed its conditioning trials, subjects were given 1 day's retraining on the licking response, followed by four light test trials and then by 20 test trials (for Groups T + 16 TL and T + 1 TL only) to the tone spread over the next 5 days. No shocks were scheduled during retraining or during test trials.

It was expected that in group T + 16 TL pretraining on T would block learning about L, i.e. that this group would be less suppressed on L test trials than group 16 TL. If

complete blocking occurred, they should show no more suppression than a group receiving T trials only. Since, however, Kamin has shown that some learning about the added component may occur on the *first* compound trial, it seemed wiser to allow for such an effect here. Accordingly, group T + 1 TL was given one compound TL trial. Comparison of the performance of group T + 16 TL with that of group T + 1 TL on L test trials would indicate whether learning about L occurred during the extra compound trials received by the former; comparison of thire performance on T test trials would indicate whether further learning about T had occurred.

Results

The results for the second pretest trial to each stimulus revealed suppression ratios not significantly different from 0·50 for all groups. Mean suppression ratios on the light test day, and over the 5 days of testing to tone are shown in Figure 1. An analysis of variance performed on suppression ratios to the light revealed a significant difference between the three groups ($F(2,27) = 4{\cdot}60$, $P < 0{\cdot}05$). Subsequent *t*-tests showed that group T + 16 TL was less suppressed than Group 16 TL ($t(18) = 3{\cdot}71$, $P < 0{\cdot}01$), but did not differ from group T + 1 TL ($t = 0{\cdot}64$). The suppression shown by group 16TL proves that the light was an effective stimulus when presented in compound with the tone. It should be noted that group T + 16 TL showed if anything less suppression to the light than did group T + 1 TL: there can be little question therefore that prior training on the tone effectively blocked learning about the light. An analysis of variance performed on the test results to tone revealed a significant decline in suppression over days ($F(4, 60) = 20{\cdot}39$, $P < 0{\cdot}001$), and a significant group effect ($F(1, 18) = 4{\cdot}91$, $P < 0{\cdot}05$). The interaction was not significant ($F < 1$).

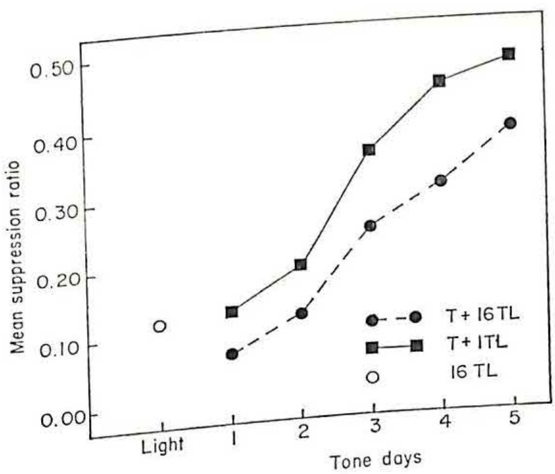

FIGURE 1. Experiment II: Mean suppression ratios on test trials to light and tone.

Although, therefore, prior training on tone caused complete blocking of the light, nevertheless the additional compound trials given to group T + 16 TL by comparison with group T + 1 TL resulted in greater suppression to the tone over a series of extinction trials.

Quarterly Journal of Experimental Psychology (1971) **23**, 126–134

HABITUATION OF CONDITIONED SUPPRESSION

J. R. MILLENSON AND JILL G. DENT†
Oxford University

Habituation of a conditioned emotional response was investigated using a procedure which eliminated contaminating temporal discriminations. Three rats were trained to bar press on a random interval 60 sec schedule of milk reinforcement and variable duration tone-shock pairings were superimposed upon this baseline. Very little recovery from conditioned suppression was found over 60 sessions of testing and no systematic differences were found after a month's "vacation" from the procedure. Analysis of responding within the CS period showed uniform suppression. The data are discussed in terms of stimulus predictability.

Introduction

One effect of a pre-aversive stimulus superimposed upon an appetitively maintained operant baseline is to suppress the frequency of that operant (Estes and Skinner, 1941). This effect known, variously as conditioned suppression, the conditioned emotional response (CER), or experimental anxiety, may be alleviated by prolonged exposure to the procedure (Hendry and Van Toller, 1965). Although this alleviation suggests a kind of sensory adaptation or habituation process, several investigators have implicated the gradual development of Pavlovian inhibition of delay as the responsible factor (Millenson and Hendry, 1967; Zielinski, 1966). Such inhibition takes the form of marked temporal discriminations that develop when fixed duration pre-aversive conditioned stimuli (CS) are used. When these temporal discriminations are fully established, baseline operant responding is typically negatively accelerated throughout the pre-aversive stimulus. Indeed, the onset of the CS produces no immediate response decrement at all, and maximal suppression to the CS is retained only in the final moments just before the presentation of the aversive unconditioned stimulus (US−) (Millenson and Hendry, 1967).

Nevertheless, the established existence of developing Pavlovian inhibition of delay does not conclusively rule out some kind of habituation process also playing a role in recovery from conditioned suppression. Casual observation confirms that autonomic fear responses such as freezing, defecation, urination, and piloerection which are initially elicited by pre-aversive stimuli tend to disappear altogether with repeated CS–US− trials. In general there is a tendency for aversive stimuli, such as electric shock, to lose their potency with repeated use (Azrin and Holz, 1966). Certainly, since Sechenov, habituation has been assumed to be a cardinal property of emotional behaviour; yet true habituation or adaptation of experimental anxiety unconfounded by inhibition of delay or learned postures

† Present address: Dept. of Psychology, Harvard University, Cambridge, Mass., 02138, U.S.A.

that actually reduce the effectiveness of the US⁻ has never been conclusively demonstrated.

The identification of an habituation process is not without practical significance, for the CER is frequently used as a supposedly stable, unchanging baseline against which various treatments such as drugs, CNS lesions and brain stimulation are studied (cf. Davis, 1968). By examining prolonged exposure to variable duration CS periods (Bersh, Schoenfield and Notterman, 1953), and the use of a very mild aversive US⁻, we sought to prevent the usual temporal discriminations and thereby expose any habituation process in conditioned suppression.

Materials and Method

Subjects

Three experimentally naive hooded rats, approximately 3 months old at the beginning of the experiment served as subjects. They were maintained at 80% of their free feeding weight throughout the experiment.

Apparatus

A single Campden Instruments operant test chamber (CI410) was used. The chamber, 24·2 cm long, 21·7 cm wide, and 20·3 cm high from the shock grid to the roof, was situated in a ventilated metal chest to mask extraneous noise. The ventilating fan gave a resting noise level of 70 dB in in the chamber. Two stainless steel levers protruded 2 cm from the front wall of the chamber, 3·2 cm above the grid floor and 10·8 cm apart. The levers required a downward force of about 5 g to close a reed switch attached to their rear extension. Only the right lever was used in the present experiment. Three 2·8 W stimulus lights, spaced 3 cm apart, were situated 12 cm above the levers.

Reinforcement, consisting of a 0·1 ml drop of condensed milk, was delivered by means of a motor driven dipper into an opening in the floor of a recessed tray situated between the two levers at floor level. The tray was covered with a lightly hinged Perspex door which, when nosed open, operated a microswitch recording tray entries. A 2·8 W lamp illuminated the tray during reinforcement delivery. Reinforcement duration (3 sec) was timed from the moment the tray microswitch was operated by the rat opening the tray door.

The CS signal consisted of a synchronous clicker and stimulus lights pulsing at 5 pulses/sec. The Campden Instruments click generator used for the CS produced a noise level of approximately 77 dB in the chamber. The US⁻ consisted of 0·5 mA, 0·5 sec, 50 Hz electric shock delivered by a Grayson-Stadler shock generator type E1064GS scrambled through 16 stainless steel rods spaced 1·27 cm apart making up the grid floor of the chamber. A PDP8/S digital computer programmed in ACT (automated contingency translator), a computer language designed for the control of psychology experiments (Millenson, in press), controlled the reinforcement contingencies and the presentation of the CS and US⁻ at random intervals. The computer also collected and printed out data from the rats' bar-pressing during a pre-CS control period and during each CS period. During the final experimental phase, the computer also collected the number of responses emitted during each successive 15 sec of the CS.

Procedure

Preparatory training

Following gradual reduction to 80% of their free feeding weights, the rats were shaped by hand to press the right lever for 3 sec access to condensed milk diluted 1 to 4 with tap water. They were then placed on continuous reinforcement for four sessions of 50 reinforcements each. Next, they were introduced to intermittent reinforcement (random interval (RI) 15 sec, minimum interreinforcement interval = 1 sec (cf. Millenson, 1963)

for one session of 50 reinforcements. Then a further 50 reinforcement session was run with a RI30 sec schedule in effect. Finally, each rat was placed on an RI 60 sec schedule of reinforcement which remained in operation for all subsequent 1 h long sessions.

Introduction of CS alone (5 *sessions*)

After ten RI 60 sec sessions all subjects exhibited a steady rate of responding. The clicker/light CS was then introduced alone every 4 min on average to verify its neutrality. RI 60 sec milk reinforcement contingencies for bar pressing in force. CS duration was variable, averaging 2 min in length, and each presentation was drawn from a rectangular distribution of 15 equidistant intervals from 15 to 225 sec. Intertrial intervals (ITI) were random in length, subject to the restriction that the minimum ITI was 2 min 15 sec, and the mean 4 min.† The minimum ITI was selected to allow a 2-min measure period before each CS.

CS-US⁻ pairings (60 *sessions*)

After 5 sessions of CS alone, shock was introduced. The procedure was exactly as above, except that the termination of the variable CS was now accompanied by a 0·5 mA, 0·5 sec shock delivered to the feet of the animal via the grid floor of the chamber.

Vacation

The rats were then placed on a free feeding schedule for one month, during which time no experimental sessions were run and the animals remained in their home cages. At the end of this time, the animals were again reduced to 80% of their new free feeding weight (about 10 g above the previous determination) and were given 2 sessions of RI 60 sec retraining, with no CS or US being presented.

Post-vacation CS-US⁻ pairings (20 *sessions*)

The CS-US⁻ pairings of the earlier phase were reinstated with identical parameter values.

Results

Figure 1 shows session-by-session CS and pre-CS lever pressing rates throughout the entire experiment. During the CS-alone trials response rates were systematically rising, but no differential stimulus effect can be discerned. With the introduction of 0·5 mA shock, all subjects showed an initial sharp reduction in response rate both inside (open circles) and outside (solid circles) the CS period. With continued exposure to the procedure, however, this suppression gradually became confined to the CS alone; the pre-CS baseline recovering to normal after about 10 to 20 shock sessions. For rats H9 and H11, the response rate in the CS remained considerably suppressed for some 30 sessions. Around session 30 all subjects showed an increase in CS response rate which quickly reached an asymptote, and there was no further systematic increase from session 30 to session 60. Rat H12 showed the highest degree of recovery and for a few sessions showed a return to pre-shock rate within the CS. However, this effect was unstable and by session 60 this subject's CS rate was less than half of the pre-CS rate.

The 1 month vacation had a consistent systematic effect only for H12 who showed a sharp drop in CS rate which was rising toward pre-vacation level when the experiment was terminated.

†The ITI length was of the form ITI $= K + x$ sec, where $K = 120$, and x was geometrically distributed with a minimum value of 15 and a mean of 120.

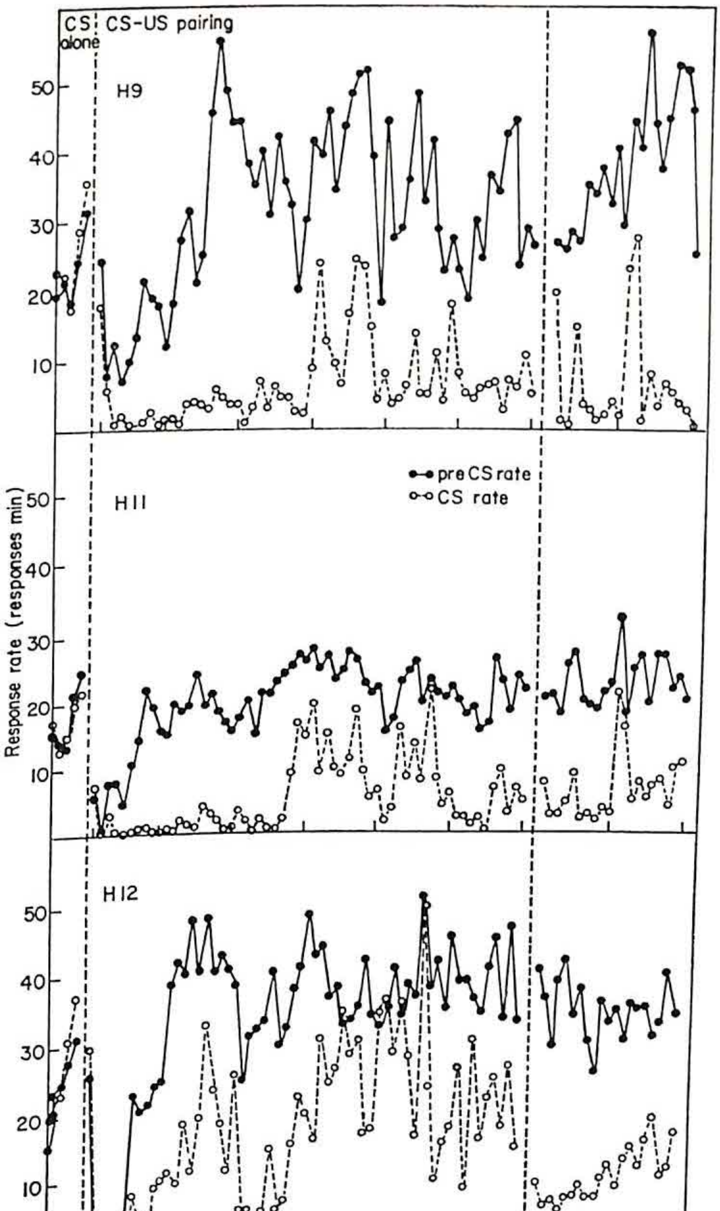

FIGURE 1. Mean daily CS and pre-CS response rates for the three subjects. The left panel shows data from the sessions of CS alone, the middle panel from the 60 sessions of CS–US pairing, and the right panel from the post-vacation session of CS–US pairings.

Conditioned suppression is conventionally reported in terms of relative CS response rate to pre-CS rate. In Figure 2, the ratio of lever response rate in CS to lever response rate in pre-CS period is plotted session-by-session for the entire experiment. These ratios show that relative suppression drops precipitously following the introduction of shock, rising only gradually by session 40 to 50 to an

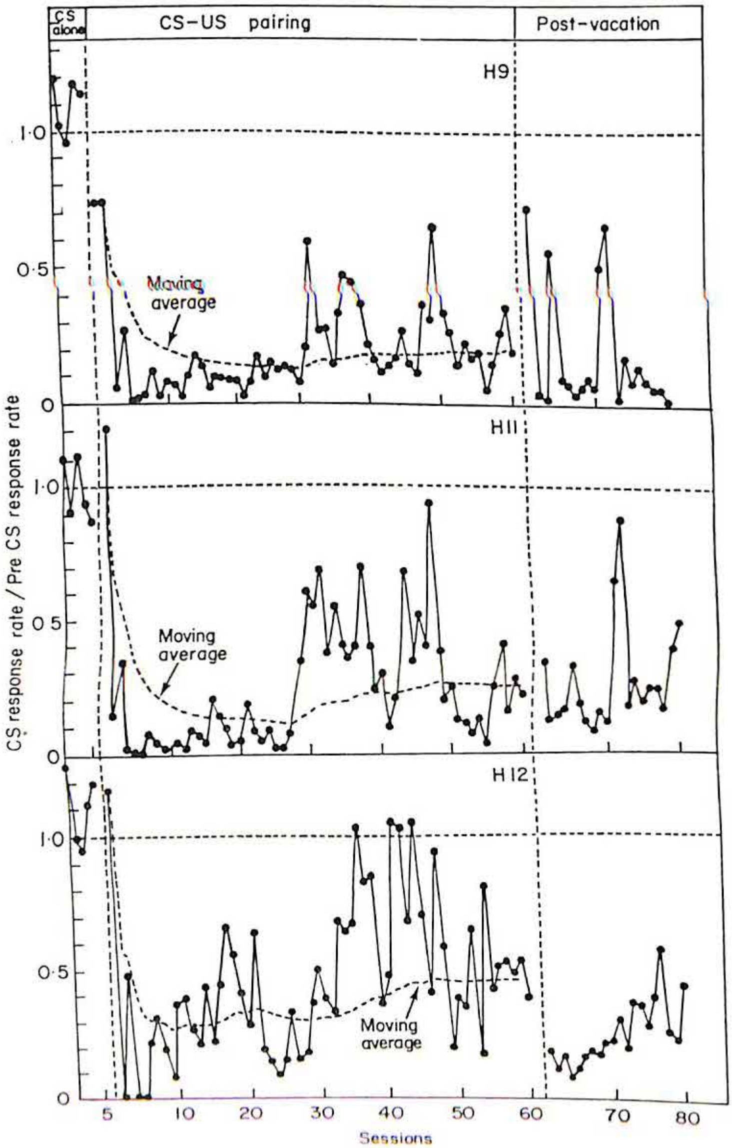

FIGURE 2. Session-by-session suppression ratios for each subject. Column designation as in Fig. 1. A ratio of 1 indicates CS rate = pre-CS rate, a ratio of 0 indicates complete suppression in the CS. The curve of moving averages shown represents means calculated for each session on the basis of all data up to and including that session.

asymptote of 0·2 to 0·4. Again, subject H12 shows the only reliable effect of the vacation: an initial drop and subsequent rise to a stable partial suppression.

Within the CS period itself, response rates were calculated for successive 15-sec periods. These within-CS rates appear as the solid circles in Figure 3 derived from the last 15 sessions of the experiment and averaged for the three subjects. The straight line with zero slope drawn through the median point represents the null hypothesis that these points are samples from a single distribution of response rates. Considering the points above and below the mean CS

duration of 120 sec, 4 out of 7 above the mean value exceed the median CS rate, while 3 out of 6 below the mean value are exceeded by the median CS rate. A sign test indicates that the probability of such a result occurring by chance is 0·71; hence the null hypothesis cannot be rejected.

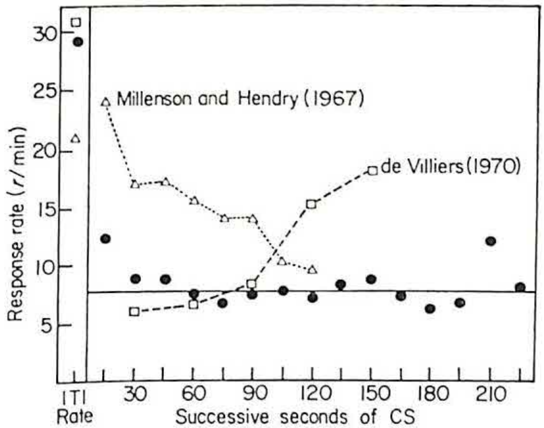

FIGURE 3. Within-CS response rates. Solid circles represent mean response rates in successive 15 sec periods of the CS from sessions 65–80. A line of zero slope is drawn through the median rate point. The dotted lines and open triangles represent comparable data of Millenson and Hendry (1967) with fixed duration CS periods. The dashed line and open squares represent comparable data of de Villiers (1970) with geometrically distributed CS durations.

Discussion

In the literature, habituation has been interpreted in three main ways: (1) As an exhaustive process, (2) an inhibitory, and (3) as anticipatory (McDaniel and White, 1966). The process is defined as a response decrement due to repeated stimulation, and is usually distinguished from fatigue in that a novel stimulus will dishabituate the response. It is generally considered to be a mechanism whereby the CNS can avoid arousal by efferent control over sensory input at several points in each receptor system. If this damping of arousal is a conditioned anticipatory response, then it would be expected that periodicity or temporal uncertainty of the presentation of the stimulus should be an important parameter, for with random presentation there are no temporal cues available for anticipation to occur. Inhibitory theories of habituation would not consider this to be an important variable, since only the rate and intensity of the stimulation are considered critical parameters for a concentration of inhibition within the CS to occur (Pavlov, 1927).

McDaniel and White (1966) tested the anticipatory hypothesis by comparing periodic versus random stimulus presentation in habituation pretraining before an escape/avoidance task employing the same stimuli. They concluded that periodicity was an important variable since cats given random habituation presentations of the stimulus showed significantly less response decrement to the stimulus in the later escape/avoidance situation. Similarly, Fox (1964) found the probability of stimulus occurrence to be a major determinant of changes in evoked brain potential amplitude.

In most CER studies CS onset is rendered unpredictable through the use of partially randomized ITI's. Nevertheless, the duration of CS itself is usually a relatively long (measured in minutes) fixed value. Hence the occurrence of the US is predictable and the conditions for anticipated habituation are guaranteed. The dotted curve in Figure 3 taken from data of Millenson and Hendry (1967), with 2 min fixed CS periods, 4 min fixed ITI's and similar shock parameters to the present experiment, shows the extent to which such fixed duration CS periods encourage differential recovery from conditioned suppression. Comparison of the form of this curve with the horizontal line describing the plotted points from the present experiment is sufficient to explain why Davis, McIntyre and Cohen (1969) found overall suppression to be greater with variable vs. fixed length warning stimuli.

Nevertheless, CS variability alone is insufficient to produce uniform suppression. When a geometric approximation to an exponential distribution of CS intervals is employed, stable suppression assumes the form of the dashed curve in Figure 3 (de Villiers, 1970). Suppression is maximal only in the early portions of the CS, abating as the interval continues without the delivery of a US^-. This result is especially noteworthy, because with an exponential distribution of intervals the probability of CS termination and US^- onset at any time t is a constant.†

Holding constant the probability of shock at any second during a fixed duration warning signal, Rescorla (1968) found a similar temporal pattern of suppression. Rescorla suggested that this effect could be due to the increased probability of reinforcement (and therefore "pressure" to respond) that occurs the longer the subject has refrained from pressing. The procedure used in the present experiment is susceptible to the same baseline "pressures", but failed to yield this non-uniformity in suppression. Evidently the suppression pattern in a CS is primarily under the control of parameters of the CS duration distribution.

One such parameter is the relative frequency of CS intervals. In a geometric distribution the relative frequency is inversely proportional to the length. As de Villiers' (1970) data in Figure 3 indicates, stable suppression is also inversely proportional to the length of the CS. Although the match is not perfect (the suppression slope is consistently flatter than the exponential distribution) the correspondence is close enough to suggest that this relative frequency plays an important role in determining relative suppression.

The present study thus adds support towards the view of McDaniel and White (1966) that extensive habituation of the CER, like other habituation, only occurs under conditions where the unconditioned stimulus is predictable, as in the fixed duration CS procedures; or predictable on the basis of relative past frequency, as in the exponential distribution of CS lengths used by de Villiers (1970). The procedure used in the present experiment provided no such temporal cues, and hence a uniform, low suppression was obtained. It is, however, perhaps surprising

† In analogous random interval operant reinforcement schedules in which the probability of a reinforcement assignment at any time t is a constant, the probability of response also tends to remain constant (Catania and Reynolds, 1968). The present finding that response suppression fails to obey a similar rule suggests a fundamental difference between either operant and respondent susceptibilities to reinforcement, or between positive and negative reinforcement.

that *any* adaption at all occurred, for Fox (1964) points out that any development of an expectancy is unlikely with random stimulation. Nevertheless, while temporal uncertainty is reduced, still the actual presentation of the CS predicts shock at *some* time, a contingency which may permit a modicum of habituation in its presence.

In the present experiment the "vacation" had small and inconsistent effects. Rachlin (1966) gave birds a vacation from shock after continuous punishment of intermittently food reinforced key pecking. When shock was reintroduced, renewed suppression was found, even though prior to vacation there had been almost complete recovery from suppression. The differences between the two vacations are different enough to discourage any systematic comparison.

The difference in recovery found in this study as compared to the punishment studies (Azrin and Holz, 1966) may again be due to the predictability of the stimulus. With punishement for every response, the stimulus is maximally predictable, and hence habituation should occur rapidly. It is also possible that the high rate of presentation of the stimulus is a factor in the punishment studies. While this might explain the fast course of habituation, it would not explain the lower asymptote reached after a series of 80 sessions with mild shock in the present study.

The present study attests to the importance of temporal cues as a critical parameter in the habituation of the CER, and without such cues, the attenuation of the emotional properties of the stimulus remains incomplete. This has wider implications for studies concerned with the interaction of various drug, drive, and schedule conditions with the CER, for in all studies using fixed CS durations the results may be confounded by the temporal discriminations fostered by these variables.

Supported in part by grant No G9696425/C from the U.K. Medical Research Council to L. Weiskrantz. We thank Peter A. de Villiers for his generous advice and assistance.

References

AZRIN, N. H. and HOLZ, W. C. (1966). Punishment. In HONIG, W. K. (Ed.), *Operant Behaviour Areas of Research and Application.* New York: Appleton-Century-Crofts.

BERSH, P. J., SCHOENFELD, W. N. and NOTTERMAN, J. M. (1953). The effect upon heart rate conditioning of randomly varying the interval between conditioned and unconditioned stimuli. *Proceedings of the National Academy of Science, U.S.A.* **39**, 563–70.

CATANIA, A. C. and REYNOLDS, G. S. (1968). A quantitative analysis of the responding maintained by interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, **11**, 327–83.

DAVIS, H. (1968). Conditioned Suppression: a survey of the literature. *Psychonomic Monograph* Supplements, **2**, 283–91 (whole no. 30).

DAVIS, H., MCINTYRE, R. W. and COHEN, S. I. (1969). Fixed and variable duration warning stimuli and conditioned suppression. *Journal of Psychology*, **73**, 19–25.

DE VILLIERS, P. A. (1970). The effects of prefeeding on conditioned suppression-evidence for a motivational interpretation of anxiety? Unpublished B.A. thesis, Oxford University.

ESTES, W. K. and SKINNER, B. F. (1941). Some quantitative properties of anxiety. *Journal of Experimental Psychology*, **29**, 390–400.

Fox, S. (1964). Evoked potential habituation rate and sensory pattern preference as determined by stimulus information. *Journal of Comparative and Physiological Psychology*, **58**, 225–32.

Hendry, D. P. and Van Toller, C. (1965). Alleviation of conditioned suppression. *Journal of Comparative and Physiological Psychology*, **49**, 458–60.

McDaniel, J. W. and White, R. K. (1966). A factorial study of the stimulus conditions of habituation. *Perceptual and Motor Skills*, **23**, 259–70.

Millenson, J. R. (1963). Random interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, **6**, 437–43.

Millenson, J. R. (1971). A general language for on-line control of psychological experimentation. *Behavior Science*, in press.

Millenson, J. R. and Hendry, D. P. (1967). Quantification of response suppression in conditioned anxiety training. *Canadian Journal of Psychology*, **21**, 242–52.

Pavlov, I. P. (1927). *Conditioned Reflexes*. London: Oxford.

Rachlin, H. (1966). Recovery of responses during mild punishment. *Journal of the Experimental Analysis of Behavior*, **9**, 251–63.

Rescorla, R. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, **66**, 3–8.

Zielinski, K. (1966). Inhibition of delay as a mechanism for the gradual weakening of the CER. *Acta biologae experimentalia* (*Warsaw*) **26**, 407–18.

Received 21 June 1970

Quarterly Journal of Experimental Psychology (1971) 23, 135–145

FOOD PREFERENCES ACQUIRED BY ASSOCIATION WITH VARIATIONS IN AMINO ACID NUTRITION

D. A. BOOTH AND P. C. SIMSON

Laboratory of Experimental Psychology, University of Sussex

Protein-deprived rats were given, on one day, a balanced mixture of amino acids followed by access to protein-free food having a distinctive odour. On another day, an imbalanced (histidine-free) amino acid mixture was given just before food having another odour. The rats afterwards preferred the balance-paired odour to the imbalance-paired odour. The preference was acquired whether the duration of odour presentation, or the amount of odourised food taken, was kept constant on the two conditioning days. Retention of the preference seemed unattenuated after 4 weeks. An attraction to the balance-paired odour (relative to odours paired with a water load) contributed to the acquired preference. There was also a relative aversion to unfamiliar odours when they had been paired with imbalance. Such acquired chemosensory control of preferences, together with an anorexigenic effect of imbalanced amino acid mixtures, can account for characteristics of feeding behaviour under conditions in which the diet is deficient in an essential amino acid.

Introduction

Loss of appetite on inadequate diets has long been recognized (Rose, 1938). One form of inadequacy is amino acid imbalance. This occurs in diets containing protein of poor biological value (proportionately low in some essential amino acid), or containing artificial mixtures of amino acids lacking an essential amino acid (that is, one which cannot be synthesised in the body). If ingestion of an amino acid imbalanced diet follows a few days of feeding on protein-free food and an overnight food deprivation, a depression of food intake occurs within a few hours—5 hr in the case of threonine deficiency in rats (Sanahuja and Rio, 1967) or 2 hr in the case of valine deficiency in chicks (Zimmerman and Scott, 1967). Harper and co-workers found in addition that, if the rat is given a choice, within a day or two it comes to take more of a diet balanced in essential amino acids than of an imbalanced diet, or even to take a protein-free diet rather than an imbalanced one (Sanahuja and Harper, 1962; Rogers, Tannous and Harper, 1969). This differential intake develops even though the only known chemical difference between the alternatives is the presence or absence of 0·13% histidine. Harper (1964) has argued that this would not be discriminable by taste in the presence of the 6% protein content, including 3·8% free amino acids, of these diets. If this assumption is correct, the differential intakes cannot be based on an innate reaction to some taste difference that generally distinguishes balanced from imbalanced diets. Nevertheless, the eventually consistently lower intake of the histidine-free diet is likely to be based on some discrimination—the alternative, that the diets are accepted equally often

as the choice for a given feeding bout and the imbalanced diet has a much more rapid satiating effect, does not seem very plausible. The position of the diet was regularly changed in at least some of these experiments. Perhaps the rat detected consistent odour differences between the diets, introduced either by mixing or storage conditions or by the rat's incidental reactions towards the end of its more lengthy feeding bouts.

Le Magnen (1956, 1969) has demonstrated that postprandial events can make arbitrary food odours acquire intake- and preference-depressant properties. It is also now established that taste-discriminated intake depressions and elevations can be reinforced by toxic reactions (Garcia and Ervin, 1968). This "toxiphobia" has recently been found to occur for odours as well as tastes (Garcia and Koelling, 1967; Pain and Booth, 1968). Such mechanisms could mediate the development of selective avoidance of, or aversion to, imbalanced diets. The present experiments were an attempt to demonstrate in rats the acquisition of such changes in chemosensory preferences when they were reinforced by accentuation or correction of amino acid imbalance. We also tried to see whether the strength of any acquired preference was correlated with the extent of immediate depression of food intake caused by the imbalanced mixture, because one of the undecided issues concerning metabolic conditioning of oral acceptability is whether the reinforcing event is identical with the satiating or anorexigenic reactions to the administered materials (Lovett and Booth, 1970).

Method

Pre-treatment

The subjects were male albino rats of a Sprague-Dawley strain supplied by Shaw's Farm, Blackthorn, Bicester, Oxfordshire. They were housed individually with free access to Spillers autoclaved Small Animals Diet for at least a week after delivery. Five days before conditioning they were placed on a protein-free diet based on Crystal Gum dextrine (Laing-National, Manchester). A Rogers-Harper salt mix (5%), corn oil (5%), vitamins and choline chloride were included. To minimize spillage, the food was presented in a small tin within a larger tin bolted to the floor of the cage. In the initial stages of protein depletion the rats were accustomed to drinking from water tubes made from 10 ml pipettes, by presenting in them 15 to 20 ml of 9% casein hydrolysate (British Drug Houses, Poole) in 2 or 3 portions. In the last 2 days before conditioning, water was available only from pipette tubes and the only nutrient was the protein-free diet.

There were two main groups in Experiment I, each of nine rats. An additional four rats were given access to odourized diets with the experimental 18 but received no conditioning loads. Body weights were initially 100 to 150 g in the first main group and 190 to 260 g in the second group, used 2 weeks later. During protein depletion 10 to 25 g and 10 to 36 g, respectively, had been lost by the first day of conditioning (mean losses 14 and 11%), an effect very similar to that observed by Kumta and Harper (1962). Twelve rats were pre-treated for Experiment II.

Conditioning

After an 18-hr fast, the rats were loaded with either a balanced or deficient amino acid mixture, or water and then given access to an odourized or plain protein-free diet for 6 hr. In the first group of Experiment I, 5-ml loads were stomach-tubed under light ether anaesthesia; in the second group, 8-ml loads were presented orally for 30 min (water being flavoured with 2mM sodium saccharine), and any remaining fluid tubed. In Experiment II the rats were intubated without anaesthesia. The concentration of the amino acid loads was 9·1 g/100 ml. The balanced amino acid mixture was based on the proportions

of L isomers given by Sanahuja and Harper (1962) for their amino acid-supplemented 5% fibrin diet: leu 1·03, ile 0·49, val 0·45, his HCl 0·26, met 0·36, thr 0·57, try 0·30, phe 0·58, lys HCl 1·15, arg HCl 0·67, cys 0·10, tyr 0·40, glu Na 3·74. The imbalanced mixture was identical except that histidine hydrochloride was omitted.

In Experiment I this conditioning treatment was given on each of three successive days so that each rat had each load condition paired once with each of three food odour conditions. The odour conditions were the usual plain protein-free diet, citral-odourized (lemongrass oil), and eucalyptus oil odourized. The oils (Griffin and George, London) were added to the diet daily just before use (4 μl/100 g). Each of the 6 trios of rats received the load conditions in a different sequence. Among the 3 trios in each of the two groups, each load was paired with a different odour condition in each of the 3 rats of a trio according to the standard Latin square. Two samples of diet were presented and water in a pipette tube. Intakes were measured every 2 hr.

In Experiment II there were two conditioning days. Six rats were tubed the imbalanced mixture on the first day and the balanced mixture the following day; the other six had the mixtures in reverse order. Half of each sub-group was presented with citral-odourized protein-free diet after their balanced load, and half with eucalyptus-odourized diet; the other odour in each case was paired with the imbalanced load. On the first conditioning day, the rats were given access to odourized food for 6 hr. On the second day, however, their access was interrupted or extended so that it was ended at the half hour nearest to the time by which they had eaten the same amount as on the previous day. Access periods on this second day ranged from 3 to 8 hr, mean 4·7 hr.

Testing

A day with 6-hr access to Small Animals Diet followed conditioning, and preference testing followed from the subsequent day. Each rat was presented with two cans of protein-free food, one on each side of the back of the cage. The choice given on the first test day was between the odour conditions which had been paired with loads of balanced and imbalanced amino acid mixtures. Second and third test days were included in Experiment I. On these, the odour conditions which had been paired with water load were tested in turn against odour conditions which had been paired with balanced and imbalanced loads—half the rats being given the choice water-odour vs. balance-odour first, and half second. Each preference test lasted 3 hr. Food and water intakes were measured hourly. An hour after testing the rats were given 1 hr's access to Small Animals Diet.

In Experiment I a retention test was made a month after conditioning in 8 surviving rats from the first main group and a fortnight after conditioning in the 9 rats from the second main group. The rats had in the meantime been maintained on Small Animals Diet. The test was a single 3-hr choice between odour conditions which had been paired with balanced and imbalanced loads.

Reversal conditioning

The 14 survivors from conditioning in Experiment I were put through the above conditioning procedure again, with changes in which odour condition followed which load. Only pairs of odour conditions were reversed for rats from the first group, 2 each being reversed water vs. balanced and water vs. imbalanced and 3 having the odours reversed between balanced and imbalanced. In the 7 rats from the second group, all 3 odour conditions were paired with new load conditions, the odour which had for a given rat been paired with balanced amino acids being paired with the imbalanced load for re-conditioning.

Results

Experiment I

Intakes immediately after loads

The two groups did not differ reliably in the immediate effects of gastric loading and so their data for this stage of the experiment were combined (Fig. 1).

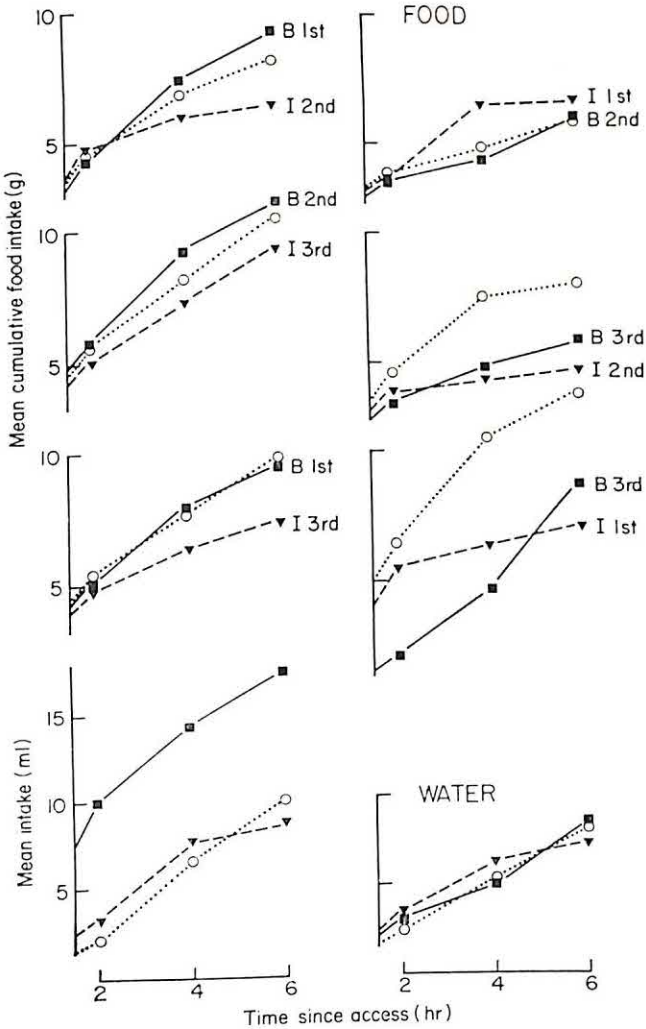

FIGURE 1. Food and water intake after intragastric loading in Experiment I. The top 3 left-hand panels give cumulative food intake means for trios of rats which received loads in sequences in which balanced amino acids preceded imbalanced, the bottom panel being the mean water intake for these 9 rats. The right-hand panels are for imbalance preceding balance. ○, After water load; ■, after balanced load; ▼, after imbalanced load.

Analyses of variance were performed on intakes for each 2-hr period and the total 6 hr, an F-ratio probability of 1% or less being considered reliable. Tukey's test revealed no non-additivity effects of probability less than 5%. The 6-hr food intakes showed reliable variance among the three types of stomach load, among the six sequences of load types (P's $<0{\cdot}001$) and among the three food odour conditions ($P<0{\cdot}01$). The variation among loads was highly reliable at 2 to 4 and 4 to 6 hr ($P<0{\cdot}001$) but not in the first 2 hr after gastric loading. There were no other main effects reliable to this degree in any particular 2-hr period, except that between odours at 2 to 4 hr ($P<0{\cdot}01$). There was in addition a reliable interaction between loads and load sequences seen over the 6 hr and at 2 to 4 and 4 to 6 hr (P's $<0{\cdot}01$).

These variations were attributable to the following effects on the basis of t-tests (compare mean data for trios of rats in Fig. 1).

Between loads of amino acids or water. Over the 6 hr and at 2 to 4 and 4 to 6 hr, food intake was depressed after administration of the imbalanced amino acid mixture, relative to intakes after either balanced amino acids or water in equivalent volume (P's$<0{\cdot}05$). Water intakes did not parallel this pattern. The only reliable difference between load conditions which appeared in water intakes was an elevation in the first 2 hr after the balanced load, relative to drinking after either water or imbalanced loads. This difference was entirely attributable to rats which received the balanced load before the imbalanced (Fig. 1).

Between sequences of loads. No reliable mean difference in food intakes at any time interval after loading was found between any pair of load sequence types. However, when the loads/load-sequences interaction was examined, it appeared that the depression of feeding after the imbalanced load varied in speed of onset and in size with the position in the sequence of three conditioning days on which the imbalanced amino acid load was administered. A greater and shorter latency effect was seen when the imbalance came second or third in the conditioning sequence than when it was first, especially if the day for balanced amino acids preceded it (contrast left with right-hand panels in Fig. 1).

Between odour conditions. The variation amongst odour conditions was attributable to smaller mean intakes of odourised food (8·37 g of lemongrass and 8·08 g of eucalyptus) than of food without added odourous oils (8·53 g). This effect was also seen in the 4 rats outside the main experiment, which were never given amino acid loads but did go through the same 3-day cycle of food odour conditions (mean intakes 10·55, 9·55 and 11·48 g respectively).

Intakes in preference tests

The difference between the two food intakes seen when the rats were given a choice of food in two odour conditions was the preference indicant used for statistical analysis. Although the reliability of an effect in the first main group of the experiment was not necessarily paralleled by a similar reliability in the second (Fig. 2), no reliable difference was found between the groups in any type of preference. The only substantial effect revealed by a combined analysis of variance was among the types of load which had been associated with the pair of odours used in this preference test ($P<0{\cdot}001$). There may also have been slight odour and load/odour interaction effects (P's$<0{\cdot}05$). Setting the acceptable level of reliability at $P<0{\cdot}01$, separate analyses for the two groups revealed the between-load-pairs effect in both groups and a load-pair/odour-pair interaction in the second group only.

Balance-paired odour vs. imbalance-paired odour. A preference for food of the odour which had been paired with the balanced amino acid mixture over the imbalance-paired odour contributed to the loads variance in the first group definitely, in the second group marginally (Fig. 2) and definitely in the groups combined ($P<0{\cdot}02$, 2-tailed correlated t-test). The same preference was also indicated indirectly by combining the data from choices relative to water-paired odour

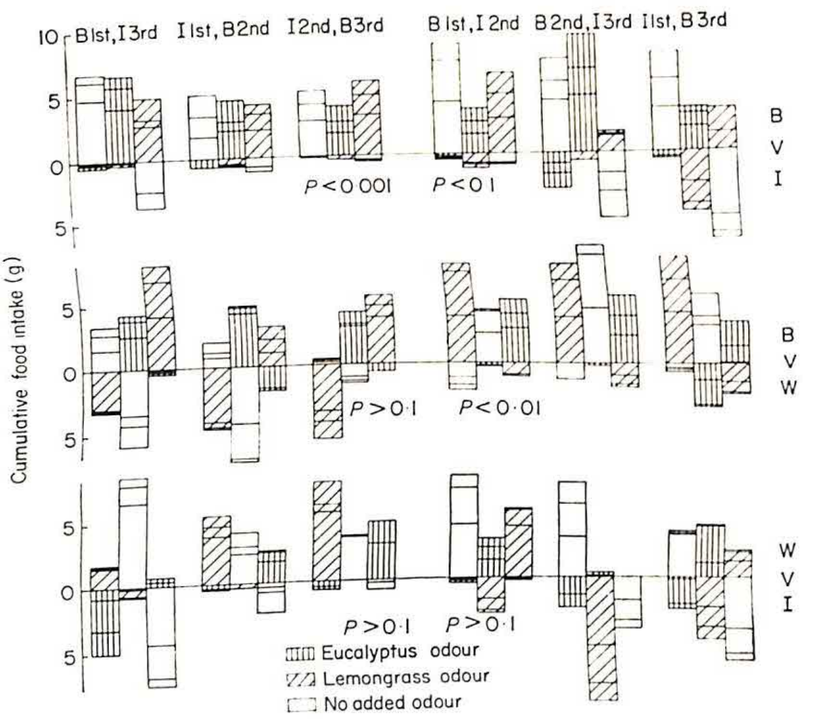

FIGURE 2. Food intakes in preference tests in Experiment I. Each column is one rat. Each row is the odour pairs associated with one pair of load conditions—B: balanced, I: imbalanced, W: water. The horizontal lines within intake columns divide them into first, second and third hours of access. The P values are for two-tailed correlated t-tests on the 3 hr total intake differences for a given association in a given group of the experiment. The first group is on the left, the second on the right of the figure.

condition: the mean differences between balance-paired and imbalance-paired odour intakes calculated in this way were 2·87 g (directly measured 4·41 g) in the first group and 5·11 g (directly measured 3·26 g) in the second, both reliable at $P<0{\cdot}05$. These direct and indirect preference measures correlated well ($r = 0{\cdot}51$, $P<0{\cdot}005$). The intake differences often represented almost complete rejection of the imbalance paired-odour (Fig. 2). The degree of preference in the choice between balance-paired and imbalance-paired odour conditions (as indicated by the difference between the two intakes) was not reliably related to the difference that had been seen between intakes at 2 to 6 hr after the balanced and imbalanced loads ($r = -0{\cdot}33$, $P>0{\cdot}1$). There was no correlation between degree of preference and the amount of increased drinking seen after the balanced load.

Balance-paired odour or imbalance-paired odour vs. water-paired odour. To get some indication of the relative contribution to the balance-odour vs. imbalance-odour preference of aversion to the imbalance-odour paired odour and attraction to the balance-paired odour, we also measured preferences relative to food of the odour paired with administration of a water load. In the choice between food of odour which had been paired with the balanced amino acid mixture, and food of water-paired odour, the mean intake difference was reliably greater than zero in group two (cf. Fig. 2), and in the combined results of the two groups ($P<0{\cdot}02$).

Thus loading with balanced amino acids increases the acceptability of an odour, especially if the amino acid mixture is wholly or partly drunk rather than stomach-tubed, which was the case for the second group of this experiment. On the other hand, neither group, nor both together, revealed a reliable preference for water-paired odour over imbalance-paired odour. However, it would be incorrect to conclude that pairing with imbalance does not introduce any aversiveness. For one thing, in the first group neither sort of preference test involving water-paired odour vs. imbalance-paired odour gave reliable results, even though the balance-paired odour vs. imbalance-paired odour gave a highly reliable preference—so there may be reduced sensitivity in tests vs. water-paired odour. For another thing an intake-depressant effect of pairing with imbalance under certain conditions is indicated on examination of the load/odour interaction, to which we now turn.

The interaction between the type of load and the odour variable used in the analysis (i.e. which pair of odour conditions followed the two amino acid loads) could be attributed to poor conditioning of reactions to the unodourized food by the imbalanced amino acid mixture. Table I gives, for all stages of both the groups,

TABLE I

Variation in strength of preference with odour condition reinforced (Experiment I)

	Mean balance-odour intake less imbalance-odour intake			
	First test	Second and third tests	Retention	Reversal
Neither load paired with plain diet	5·75 (6) †	7·95 (6) ‡	5·40 (6) §	1·88 (4)
Balanced load paired with plain diet	4·35 (6) §	0·75 (6)	3·33 (6)	4·50 (5)
Imbalance paired with plain diet	1·40 (6)	2·83 (6)	−2·23 (4)	−4·12 (5)
P of *F*-ratio	<0·01	N.R.	N.R.	N.R.

Two-tailed probability that mean does not differ from zero by correlated *t*-test: † <0·001; ‡ <0·01; § <0·05.
Number of rats per mean is given in parentheses. N.R.: not reliable (P >0·05).

the means of the intake differences between odour conditions which had been paired with balance and imbalance, divided out according to which type of load had been given with unodourized food. Reliable and large differences were obtained only when artificial odour had been added to the food paired with the imbalanced amino acid mixture, whether or not the balanced load was paired with an added odour. Hence a conditioned aversion reinforced by the imbalanced mixture must be contributing to the preference for balance-paired odour conditions over imbalance-paired.

Retention test intakes

Table II gives the intake difference between food having the balance-paired odour and food having the imbalance-paired odour on all the various occasions this preference measure was obtained, for the 13 rats which survived retention and reversal tests. A correlated t-test showed no reliable decrease in intake difference from the first preference test to the retention test.

TABLE II

Strength of conditioned preferences initially and in retention and reversal tests (Experiment I)

	Mean balance-odour intake less imbalance-odour intake (g)			
	First test	Second and third tests	Retention	Reversal
Mean	4·27	3·22	3·37	0·60
S.D.	3·25	4·30	4·25	5·84
P	<0·001	<0·05	<0·05	N.R.

The mean intake difference in the retention test was reliably greater than zero in the first group of the experiment taken by itself ($P<0·01$) and was not reliably different from that in the original test: thus retention was virtually perfect over 4 weeks by this measure. The retention score of the experiment's second main group taken alone was not reliable but as the group's initial preferences were marginal this has little significance.

Reversal

A systematic analysis of the effects of differences in type of load on immediate food intakes during reversal conditioning was not possible, because groups were incomplete and not counterbalanced. The only reliable intake differences were depressions at 4 to 6 hr after the imbalanced load, relative to intakes after balanced amino acids or water (P's $<0·01$).

The subsequent preference tests gave no reliable mean intake differences, in either group, or for all rats combined (Table II).

Experiment II

Experiment I involved a fixed time of access to odourized food after administration of balanced and imbalanced amino acid mixtures. After the imbalanced load, intake rate was depressed. It was therefore conceivable that the acquired preference merely reflected a difference in the amounts of food of the two odours which had been ingested by the rat previously to the preference test. Rather than the amino acid loads generating reinforcing events which shift preference, their effects on appetite might, for example, induce differential extinction of neophobia. To exclude such possibilities, in Experiment II the amount of odourized food eaten after the balanced load was matched with that taken after imbalance on acquisition days to within 0·4 g and did not differ appreciably on average over the group (Table III, column 2). Nevertheless a reliable preference appeared on the

TABLE III

Conditioning of preference with equalized acquisition intakes (Experiment II)

	Acquisition		Preference testing: Balance-odour intake minus imbalance-odour intake			
	Mean daily intake	Balance-day intake minus imbalance-day intake	First hour	Second hour	Third hour	Total
Mean (g)	5·99	0·025	1·66	0·73	1·18	3·57
S.D. (g)	2·58	0·24	1·62	0·73	1·18	2·85
P	—	>0·1	<0·01	<0·01	<0·05	<0·001

Two-tailed correlated t-tests, $N = 12$.

choice day for the odour which had been presented after tubing the balanced mixture of amino acids (Table III). Neither the sequence in which the amino acid mixtures were tubed nor which odour was paired with which mixture produced any reliable differences in the strength of preference for balance-paired odour over imbalance-paired odour in this experiment.

Discussion

Our results add to the evidence that a diet deficient in an essential amino acid causes loss of appetite. By rapid oral administration of a histidine-free amino acid mixture to rats losing weight on a protein-free diet, we have obtained a food intake decrement at 2 to 4 hr, continuing up to at least 6 hr, without additional imbalanced loads. The depression did not extend to water intake, which suggests that the amino acid deficiency can cause a loss of appetite specifically for food, rather than a general malaise.

We found that the intake decrement when the imbalanced mixture was given on a day subsequent to the day on which the balanced mixture had been given was greater than when the two types of mixtures were given in the reverse sequence. This does not necessarily indicate that dietary protein deficiency is unnecessary to obtain an anorexia after a single imbalanced load: the amino acids in the first administered mixture may stimulate an increase in the levels of hepatic enzymes which break down amino acids and so make the histidine-free mixture when given second produce an imbalance in the rat even more effectively than when it is given first.

Our main new finding is that temporal conjunction of amino acid loads and ingestion of food having a distinctive odour can alter the preference for food of that odour. We assume that the acquired preference is olfactory because the proportion of odorous oil was too minute (0·004%) to affect the texture, colour or taste of the food. Acquisition of the new preference was not contingent on the lesser intake of the odour which followed an imbalanced load when a fixed time of access was given. Subsequent expression of the acquired preference was not dependent on the rat being in a protein-depleted state. The preference was not appreciably extinguished during access to odourized foods for 12 hr distributed

over 3 days of preference testing without further reinforcement. Even after 4 weeks on standard maintenance food, the preference was retained without apparent loss.

The addition of a novel and quite intense odour to the protein-free food associated with imbalance seemed to be necessary to get reliable preference acquisition. This may have been a matter of discriminability against a background of several days on the plain protein-free diet, or it may be an instance of the potentiation of toxiphobia acquisition by cue novelty (McLaurin, Farley, Scarborough and Rawlins, 1964). Our failure to obtain preference reversal might derive from reduced novelty of the odours, although other factors could have been operating, such as confusion, greater age of the rats and diminished biochemical effects of the amino acid loads.

There was no sign of any positive correlation between the amount of food intake depression caused by the imbalanced load (nor indeed water intake elevation caused by the balanced load) and the effectiveness of either load as a reinforcer as indicated by the various types of preference. Thus our results failed to provide any support for the idea that the short-term effect on intake is mediated by the same biochemical mechanism as that underlying the reinforcement of preference acquisition. The mechanism is obscure in either case. It has proved difficult to define the biochemical consequences of dietary amino acid imbalance in relation to its effects on food intake (Peng *et al.*, 1969; Peng and Harper, 1969). The present findings may indicate one reason for this difficulty. An acquired preference persists long after the biochemical event which established it and is likely to change intake even when only a single food is presented. Thus biochemical abnormalities are not likely to correlate well with changes in simple measures of feeding behaviour.

The present experiments do not apportion the effective reinforcement between the balanced and imbalanced loads with any precision. One experiment gave direct evidence that association with a balanced load induces an increased acceptance of an odour and gave an indirect suggestion that the imbalanced load can decrease acceptance. Another experiment failed to resolve any shifts in balance-paired odour acceptance or imbalance-paired odour acceptance relative to water-paired odour acceptance. Subsequent work has shown that an imbalanced load can establish a strong aversion relative to a saline load in association with odourized protein-free food (Simson and Booth, in preparation).

A conjunction of the anorexigenic and the negatively reinforcing effects of dietary amino acid imbalance would explain the known characteristics of the feeding behaviour of rats in this condition. The imbalance should then not only cause a possibly unselective decrease in food intake within a few hours. It should also produce, perhaps within a few hours, selective decreases in the intake of discriminable foods in proportion to their imbalance-inducing effects. Thus an animal would become more averse to food containing grossly imbalanced protein than even to food which was completely lacking in protein (Rogers *et al.*, 1969). It could have been of considerable selective value to have a capacity to acquire a long-lasting preference for more adequate protein, as this affects intake both selectively and immediately when previously experienced food is sampled, unlike a propensity merely to reject all food, adequate in amino acids or not, which

develops with a considerable delay after ingestion of poor quality protein. The conditioning mechanism provides very long-term behavioural regulation at least for large disturbances in the supply of essential amino acids.

This work was supported by a grant from the Medical Research Council.

References

GARCIA, J. and ERVIN, F. R. (1968). Gustatory-visceral and telereceptor-cutaneous conditioning—adaptation in internal and external milieus. *Communications in Behavioral Biology*, **1**, 389–415.

GARCIA, J. and KOELLING, R. A. (1967). A comparison of aversions produced by X-rays, drugs and toxins. *Radiation Research Suppl.* **7**, 439–50.

HARPER, A. E. (1964). Amino acid toxicities and imbalances. In MUNRO, H. N. and ALLISON, J. B. (Eds.), *Mammalian Protein Metabolism*, Vol. 2, pp. 87–134. New York: Academic Press.

KUMTA, U. [illegible] and HARPER, A. E. (1962). Amino acid balance and imbalance, IX. Effect of amino acid imbalance on blood amino acid pattern. *Proceedings of the Society for Experimental Biology and Medicine*, **110**, 512–7.

LE MAGNEN, J. (1956). Le rôle des stimulations olfacto-gustatives dans les mécanismes de regulation de la prise alimentaire. *Annales de la nutrition et de l'alimentation*, **10**, 153–88.

LE MAGNEN, J. (1969). Peripheral and systematic actions of food in the caloric regulation of intake. *Annals of the New York Academy of Sciences*, **157**, *Art.* 2, 1126–56.

LOVETT, D. and BOOTH, D. A. (1970). Four effects of exogenous insulin on food intake. *Quarterly Journal of Experimental Psychology*, **22**, 406–419.

MCLAURIN, W. A., FARLEY, J. A., SCARBOROUGH, B. B. and RAWLINS, T. D. (1964). Post-irradiation saccharin avoidance with non-coincident stimuli. *Psychological Reports*, **14**, 507–12.

PAIN, J. F. and BOOTH, D. A. (1968). Toxiphobia for odors. *Psychonomic Science*, **10**, 363–4.

PENG, Y., BENEVENGA, N. J. and HARPER, A. E. (1969). Amino acid imbalance and food intake: effect of previous diet on plasma amino acids. *American Journal of Physiology*, **216**, 1020–5.

PENG, Y. and HARPER, A. E. (1969). Amino acid imbalance and food intake: effect of amino acid infusions on plasma amino acids. *American Journal of Physiology*, **217**, 1441–5.

ROGERS, Q. R., TANNOUS, R. I. and HARPER, A. E. (1969). Effect of excess leucine on growth and food selection. *Journal of Nutrition*, **91**, 561–72.

ROSE, W. C. (1938). The nutritive significance of the amino acids. *Physiological Reviews*, **18**, 109–36.

SANAHUJA, J. C. and HARPER, A. E. (1962). Effect of amino acid imbalance on food intake and preference. *American Journal of Physiology*, **202**, 165–70.

SANAHUJA, J. C. and RIO, M. E. (1967). Initial effects of amino acid imbalance in the rat. *Journal of Nutrition*, **91**, 407–15.

ZIMMERMAN, R. A. and SCOTT, H. M. (1967). Effect of fasting and feeding a non-protein diet on plasma amino acid levels in chicks. *Journal of Nutrition*, **91**, 503–7.

Received 5 *September* 1970

BOOK REVIEWS

Cognition and the Development of Language. Edited by John R. Hayes. New York and London: Wiley. 1970. Pp. + 370. £5·25.

This collection of papers is based on a symposium devoted to how children acquire language, and held in April 1968 at the Carnegie-Mellon University. A relatively small number of investigators were invited, and they have contributed seven substantial essays and two shorter chapters of comments. As is the way with symposia, few people are likely to find all the contributions of equal interest, but their range and variety provide something for everyone from the phonetician to the Piagetian. The emphasis, however, falls on syntax. This places the book nicely into its context of immediate history. It is the product of a time of growing disenchantment with purely syntactic investigations of children's speech.

Why the disenchantment with syntax? The contributors between them construct the complete answer to this question. The initial impetus to psycholinguistics was provided by the idea that psychological complexity and linguistic complexity might be directly related to one another. Psychological complexity was assessed by determining how difficult an utterance was to understand or to remember. Linguistic complexity was assessed by reference to Chomsky's theory of syntax, and, in particular, to the number of grammatical transformations involved in the formal derivation of the utterance. Unfortunately, it turned out that, after some early promising results, there was no consistent relation between the two sorts of complexity. This was the fundamental dilemma which faced the participants of the symposium, and it is interesting to see how they reacted to it.

Roger Brown and Camille Hanlon, in the first paper of the book, argue that although transformational complexity is a poor guide to adult performance, it may yet predict the order in which the child acquires his syntax. They consider tag-questions, i.e. questions of the form, "John had a nice time, *didn't he*?" Underlying the apparently simple tag, there is a whole set of grammatical transformations governing such matters as the selection of the pronoun, the formation of the negative question, and the deletion of the predicate. The evidence suggests that they are acquired in the order of their linguistic complexity. However, the authors admit that the correlation may not hold for other types of construction. This seems likely. In another paper, Carlota Smith describes a study in which children had to repeat a series of grammatical and ungrammatical sentences spoken by the experimenter. Difficulties in this task were *not* related to transformational complexity but to the sheer density of information within a sentence.

The limitations of the syntactic approach are reinforced by two further papers. Susan Ervin-Tripp investigated the way in which children learn to answer questions in an appropriate fashion. It is clear from her account that semantic factors play a much more crucial role than syntactic factors. Margaret Donaldson and Roger Wales describe a curious and persistent error in children's understanding of relational terms. They evidently pass through a phase in which they consider *less* to be synonymous with *more*, and make analogous mistakes with other relational terms. Once again, such phenomena are not likely to be explained in syntactic terms; and, indeed, Clark sketches a simple but plausible semantic explanation in his comments on the paper. One implication of all these papers is clear: developmental psycholinguistics must come to an understanding with understanding.

There remains the dilemma of linguistic and psycholinguistic complexity. Tom Bever contributes an ambitious and stimulating attempt to resolve it. He points out that linguistic performance, in common with other cognitive skills, involves three distinct levels of ability. An individual possesses a basic linguistic capacity which is evident in his grasp of the fact that words refer to objects and actions, and in his knowledge of the fundamental concepts of "actor, action and object". However, before he can appreciate that words refer to things,

he must obviously be able to identify words. Hayes and Clark, in their contribution to the proceedings, show that adults can segment the sounds of an artificial language by detecting clusters that consistently recur in unbroken sequence. Perhaps this capacity to detect regularities is the most basic of all. Bever goes on to postulate a second level of linguistic ability. The child gradually builds up a series of strategies for speaking, and a series of perceptual strategies which aid him in understanding sentences. A typical perceptual strategy is to assume that the first "noun-verb-noun" sequence in an utterance is the main clause, unless the verb is specifically marked as subordinate. Such strategies are obviously not foolproof—they are built up by experience with the language, and have no immediate relation with the sort of rules postulated by the linguist. What they do account for are the differing degrees of psychological complexity in utterances. Where they work, the utterance is simple; where they fail to work, the utterance is complex. The third component of linguistic ability is epistemological in nature. The mature individual is able to make systematic generalizations of his own about language. He can appreciate, for example, that active and passive sentences are structurally related. Grammar, says Bever, is ultimately founded upon these intuitions. It aims to account for them, rather than for actual linguistic performance. Essentially the same point is made by William C. Watt in a long and technical paper contributed after the conference. He postulates in place of Bever's strategies an alternative grammar which sacrifices certain linguistic generalizations in order to predict psychological complexity. The generalizations are not lost entirely, since they are re-captured by invoking (like Bever) a separate component of adult competence.

The idea is an attractive one. We use one set of principles (or strategies) when we speak or listen, and another set of principles when we reflect upon our language. But where does this reflective competence come from? One possibility which Bever explores is that it is an extension of the perceptual strategies. However if it were just a simple consequence of the cognitive skills we acquire and exercise in handling language, the divergence between linguistics and psycholinguistics would not have arisen in the first place. It seems that the dilemma is not resolved by building it into the human being, and that we shall have to wait a little longer for its resolution.

There are some important papers in this volume, and it certainly ought to be glanced at by everyone who is interested in the psychology of language. But it is a book for the specialist. Despite the editor's introductory survey, a number of the chapters are likely to be heavy going for the reader who has not kept up with the frightening pace of developments in the area. By the time he has mastered them, they may be obsolete.

P. N. JOHNSON-LAIRD

Optics, Paintings and Photography. By M. H. Pirenne. Cambridge: Cambridge University Press. 1970. Pp. 192 + xxiv. £4·00.

This is an interesting and well-written book. It describes how we see objects and represent them in pictures, from the standpoint of geometrical and physiological optics. The first part includes a short history of optics and of perspective theory, from Euclid, Leonardo and Kepler, through Smith (1738), La Gournerie (1959), to Helmholtz. This provides a painless and entertaining way to brush up one's optics.

It includes some basic physiological optics: Pirenne's own photographs of retinal images, taken from the outside of the sclera of an albino rabbit's eye: astronomical refraction: why the sun's rays appear to diverge from the sun, and so on.

The second part is about how we see a painting as a picture. Pirenne describes in detail a number of trompe l'oeil works of art, such as Borromini's arcade built in false perspective, so that it seems longer when seen from one end than the other, and Pozzo's enormous perspective painted ceiling in the Church of S. Ignazio in Rome, which gives the spectator the impression of a huge piece of three-dimensional coloured sculpture. This ceiling shows curious perceptual rubbery distortion when the spectator moves away from the correct viewing position in the centre of the church floor. These deformations are exactly what one would expect from the theory of linear perspective, and Pozzo himself predicted them. What is strange is that we do not usually see such distortions when we

look at an ordinary painting: according to the theory of perspective, we ought to. Pirenne argues that in looking at Pozzo's ceiling, we do not perceive the painted hemicylindrical surface, which is far up, and instead receive a stimulus corresponding almost perfectly to the objects represented: so that indeed we perceive such three-dimensional objects. When we move, the retinal image changes, corresponding to three-dimensional objects which become curved and alter their positions: and this is what we perceive. In viewing an ordinary picture, however, we remain aware of the flat canvas or paper, as well as the objects depicted: we see two geometries simultaneously, not one. It has a mixture of depth and flatness. The illusion is not perfect, so that at skew angles of viewing we are able to select the particular "normalized" appearance of the objects that we see from the correct centre of projection. Pirenne supports his position with experiments. He photographed a scene several times with a pinhole camera, with the film plane tilted to different angles but without moving the camera. If the eye is placed effectively at the camera's pinhole, different photographs will provide the same retinal image; according to conventional theories of perspective, they should all look the same, but they do not because the observer is aware of the surface of the photographs. Under suitable reduction screen conditions, of course, they can be made to look the same. It is only if special reduction-screen precautions are taken to prevent any awareness of the picture surface that all the pictures look alike and give a three-dimensional appearance.

One would scarcely guess from this book that there had ever been disagreements over the nature of perspective. Pirenne studiously avoids going into controversies, and attempts "to ascertain what is right, not who is right". This will be found disappointing or reassuring according to the reader's tastes.

The book is written in a clear, easy style. The digressions are as interesting as the main argument, and there are plenty of anecdotes and curious facts which many readers may not know—for instance, that if the lens is removed from a camera, an image just as good can be formed through the resulting aperture without the lens, if the length of the camera be sufficiently increased; or that the image of a straight line projected by a pinhole camera on to a plane surface will always be a straight line, but the image of a sphere will always be elliptical, with the long axis of the ellipse directed towards the principal point of the perspective. (The single exception is a sphere whose image is centred at the principal point.) There are about a hundred illustrations, to make all this clear, including a large number of photographs taken with a pinhole camera to avoid optical distortion and give true perspective projection. This is in no sense a psychological text, but students and researchers in perception will find it stimulating and enjoyable.

S. M. Anstis

Punishment and Aversive Behavior. Edited by B. A. Campbell and R. M. Church. New York: Appleton-Century-Crofts. 1969. Pp. 550. $11·55.

The 1960's saw a resurgence of experimental interest in the effects of punishment on behaviour. In this volume, based on a conference in the spring of 1967 at Princeton, 15 contributions toward the study of aversive behaviour are grouped into five categories: (1) Quantification of punishment, (2) Suppressive effects of punishment, (3) The conditioned emotional response, (4) Implications of Pavlovian conditioning and instrumental learning for punishment, and (5) Paradoxical effects of punishment. In addition, a brief (24 pages) summary of the discussion at the conference forms an epilogue, and E. E. Boe has appended a comprehensive bibliography of experimental papers on punishment published up to June 1967.

Although all of the data presented are obtained from animal studies, and practically the only aversive stimulus used is electric shock, still a great diversity of viewpoints, methodologies, and problems are represented. Providing a coherent conceptual integument would probably have been an insurmountable task, and the editors (wisely) did not attempt it. The five part category classification is a weak ordering, but even it fails to nest Wagner's paper on increases in general activity elicited by extinction, Kamin's contribution which describes blocking effects in Pavlovian compound conditioning due to prior conditioning

with the elements of the compound, and Mowrer's casual and culture-bound essay on unauthorized avoidance behaviour from social contingencies as the basis of human neuroses.

Nevertheless, despite the evident conceptual unevenness of the volume, its appearance will be welcomed by investigators working in the field of aversive control, and by students of punishment generally. The reasons are several. (1) A number of the papers make important contributions to our conceptualization of aversive control. (2) The discursive style permitted the authors has in a number of cases allowed comprehensive surveys and integration of general aspects of punishment. (3) The instructor in undergraduate learning and motivation courses can now, for the first time, turn to one single source as a lead-in point to a significant subset of the current live issues in the field of animal punishment and aversively maintained behaviour.

The research worker will be pleased to find that Campbell and Masterson have summarized an impressive programme devoted to establishing a number of important psychophysical properties of electric shock stimulation. The authors have made it possible for the reader to compare the aversiveness of different shock sources, and have provided rational criteria for the choice of optimal stimulators based on the number of JND's between aversion detection thresholds and tetanization thresholds. Some basis for generality is given in the use of several independent behavioural measures to test equal aversiveness of different stimulators.

Estes has provided a far-reaching reformulation of his theory of punishment based on reciprocal motivational interactions of positive and negative drive systems. Like his former interpretation, punishment is viewed as a manifestation of a response-cued conditioned emotional response (CER), but in the new theory suppression is viewed as a decrement in positive motivation. Logan's experiments, which represent a successful attempt to quantify positive incentive by titrating it against punishment, provide some indirect support for Estes' position, and also represent an important advance in the direction of the eventual scaling of positive and negative reinforcers.

Rachlin and Herrnstein provide data which they argue support a primary negative Law of Effect description of punishment. They find that the reinforcing value of an aversive situation is principally computed from the classical conditioning parameters (intensity, duration, and frequency of the reinforcers) and that any response contingencies operative are important only insofar as they affect these parameters. One consequence of this position is that CER suppression must be attributed to fortuitous punishment, a view which contrasts sharply with Estes and Mowrer who would conversely explain punishment as a special case of CER. The argument is fundamental: namely, whether punishment is to be given axiomatic status in behaviour theory, or whether it is to be derived from other principles and processes, for instance, avoidance of response-produced conditioned aversive stimuli.

Church presents a rare critical discussion of choice of measure for assessing the degree of free operant suppression that a punishment procedure produces. He concludes that a relative response decrement accounts for more of the variance due to punishment treatment than other measures, but he fails to include a discussion of the theoretical import of the various possible measures. Church's paper is notable in clarifying the role of hysteresis effects which determine whether prior exposure to shock will produce adaptation or sensitization to later punishment.

Stimulus generalization of aversive control is independently treated by Hearst and by Hoffman. The former reverses an earlier commitment to the view that aversive gradients appear flatter than appetetive gradients, and now argues that the disparities were procedural in nature, and that "fundamental differences do not seem to exist between gradients for approach and avoidance". Hoffman's principal contribution is the discovery of a peak shift in discriminative CER training, which he seeks to derive by combining hypothetical gradients of excitation, inhibition, and conditioned disinhibition. The hypothetical gradients contain far too many degrees of freedom to be convincing; moreover, when Hoffman's actual empirical estimates of the gradients are combined, a peak shift fails to fall out of the derivation.

The continuing power of Pavlovian concepts to stimulate new experiments and to integrate disparate findings appears nowhere stronger than in Maier, Seligman, and Solomon's provocative paper on fear conditioning. Parallelisms are found between Pavlovian processes in salivary conditioning and in the control of instrumental avoidance behaviour by independently established Pavlovian conditioned stimuli. Maier, Seligman and Solomon (and Miller and Weiss independently) find that avoidance training can be hampered by prior Pavlovian conditioning, especially if escape from the unconditioned shock stimulus is prevented. Maier, Seligman, and Solomon seem content to label this "learned helplessness", but Miller and Weiss suggest an analysis in terms of the active and passive behaviours that a subject may acquire in such situations and which can either interfere with or facilitate later avoidance acquisition, depending on the task. Miller and Weiss show, moreover, that when an animal can control (by escape or avoidance) an aversive unconditioned stimulus, physiological stress symptoms appear attenuated.

Seward discusses the possible enhancement of stress by conflict in several situations: jumping-stand discriminations, punishment of consummatory behaviour, and approach-avoidance conflict, but is forced to conclude that the evidence for such additional enhancement is fragmentary. The problem is a difficult one to come to grips with because nearly every conceivable punishment situation involves conflict.

The remaining papers (by Fowler and Wischner on shocking correct responses in T-mazes, and Brown on the common failure of foot shock to deter previously learned running behaviour in runways) seem to lack the generality and cogency of the other papers, and to be limited in the degree to which general conclusions can be drawn from them to go beyond the specific situations studied.

On balance the volume is a success. While the contributors have probably solved few, if any, of the fundamental problems of aversive control, they have in many cases provided critical discussions and illuminating new data. The reader who is prepared to build his own bridges between these articles will obtain access to methods for scaling of positive and negative incentive, measures of response suppression, evidence for and against Thorndike's Law of Effect, information on the relative effectiveness of contingent vs. non-contingent aversive stimulation, and the role of classical conditioning processes in instrumental punishment. That reader too will soon be aware of some of the deficiencies of the book: a too frequent obscuring of behavioural effects by averaging from artificially created "groups" of subjects at the expense of individual subject analysis, the limited generality of many of the studies which sometimes seem more concerned with the analysis of artificial properties of the apparatus than with principles of behaviour, a gross neglect of response induction processes and differential susceptibility of different classes of behaviour to negative reinforcement, and the far too ubiquitous equation of aversive stimulus and electric shock.

Probably only a reviewer will want to read every paper in such a heterogeneous collection but for the few who indulge themselves, a striking dichotomy in the kinds of theories guiding the contributors stands out clearly. Perhaps even more remarkable, the dichotomy cuts both across and within contributors. The effect of new findings is to encourage some men at some times to turn either below the skin or into the head for metaphors with which to accommodate conceptually the new results. On the other hand these same men at other times are stimulated by new findings to search out novel independent variable patterns and causal arrangements, the result of which is to reorder and reorganize the phenomena in the framework of a systematic psychological theory. Thus, as an example of the former, Maier, Seligman and Solomon are led to resuscitate the empty concept that animals learn "expectancies about the outcomes of their acts" to subsume their failure to find normal avoidance acquisition after a history of unavoidable shock. No more informative are their loose notions of incentive and emotional exhaustion used as hypothetical causes of behaviour. Miller and Weiss' notion of physiological depression seems at first glance more respectable, but it too turns out to have few, if any definable properties, and how it interfaces with behaviour is left unclear. Kamin's trendy computer analogue (scanning of memory banks) purports to "explain" blocking, but again few constraints are placed on the properties of the model, and there is a grave danger that the interest in

the model might obscure continued exploration of the necessary and sufficient environmental conditions for the phenomenon which will eventually provide a legitimate *psychological* explanation of it. The same danger is present in Estes' hypothetical "amplifier elements" whose properties remain unspecified.

Above and beyond these excursions into explanatory fiction, there are several trenhant creations of theory of another sort, in which new variables are offered as high level generalizations to subsume new effects, or as the bases for conceptual reorganization of subfields. Thus Miller and Weiss introduce an important neglected variable—the amount of information about the environment provided by the reinforcing operation against a base noise level—as a plausible explanation for learned helplessness. The proposal is based on Rescorla's distinction between contiguity and contingency (a useful map relating the two is provided by Church) and suggests a realm of experimental manipulations which, in reducing learned helplessness to a special case of insufficient reinforcement, go well beyond it conceptually. Estes' new theory of CER is another instance of a productive theoretical reordering. By reiterating and expanding the view that the CER is a drive phenomenon, many new experiments are suggested, and a number of disparate phenomena from a host of diverse studies become at once interrelated and consistent. At a somewhat less lofty theoretical level, Miller and Weiss' suggestion that a coping response reduces stress and Seward's notion that conflict increases stress are two further potentially fruitful theoretical notions whose implications should lead us to new understanding of aversive behaviour.

No review would be complete without noting the continued pervasive influence of Pavlov upon this field. Although only the Maier, Seligman and Solomon paper explicitly is devoted to Pavlovian interactions with instrumental learning, the seminal concepts the Russian physiologist introduced a half century ago abound throughout in the form of gradients of generalization, inhibitory mechanisms, classical conditioned fear and incentive as hypothetical controls of learned behaviour, compound conditioning and overshadowing, and the general commitment to underlying associative processesas plausible explanations for observed behavioural effects of shock. If this volume is a sign, it looks as though Pavlovian mechanisms, and more generally two-factor learning theory, will continue to dominate the field of aversive control for some time to come.

J. R. MILLENSON

Aphasiology and Other Aspects of Language. By MacDonald Critchley. London: Edward Arnold (Publishers) Ltd. 1970. Pp. ix + 405. £9·00.

This is a collection of Dr Critchley's writings on aphasia and kindred topics over the past 30 years. Most of the papers have been published before but several have been extensively revised and not a few virtually re-written. Although it is in the clinical study of aphasia the Dr Critchley's interests and expertise predominantly lie, he shows a remarkable acquaintance with many aspects of language and linguistics that are ordinarily considered to lie well outside the orbit of medicine. Among these are the nature of animal communication, the development of speech in the child and the relations between thought and language. He is also enlightening on the history of aphasia.

While experimental psychologists may feel that their own discipline (which has after all made contributions of some importance to the analysis of language disorder) receives less than its due, they will undoubtedly learn much from Dr Critchley's brilliant clinical and historical studies. They may also come to see avenues of approach to the problems of aphasia less dour than those of contemporary psycholinguistics. If precision of method can be tempered by imagination the future of this discipline should be bright indeed.

O. L. ZANGWILL

Explanation in the Behavioural Sciences. Edited by R. Borger and F. Cioffi. London: Cambridge University Press. 1970. Pp. 515. £5·00.

This book is mainly philosophical in tenor. Much of it is lucid and considered; some is confusing in a way which is very revealing of the state of much of "behavioural science".

The book is a compendium of articles with comments and rejoinders. The contributors are S. Toulmin on *Reasons and Causes*, C. Taylor on *Purposive Behaviour*, N. S. Sutherland on *Is the Brain a Physical System?*, D. W. Hamlyn on *Conditioning*, J. Watkins on *Imperfect Rationality*, I. C. Jarvie on *Sociology and Anthropology*, J. O. Wisdom on *Situational Individualism*, G. C. Homans on *Psychology and Social Phenomena*, R. A. Boakes and M. S. Halliday on *Skinnerianism*, H. J. Eysenck on *Personality*, N. Chomsky on *Linguistics* and F. Cioffi on *Freud*, with critiques of these made by R. S. Peters, R. Borger, J. H. Grundy, A. J. Watson, Alan Donagan, Peter Winch, Robert Brown, P. M. Blau, Karl Pribram, D. Bannister, Max Black and B. A. Farrell, respectively. Rejoinders are made by the authors of the articles.

The most outstanding interchange is that between N. Chomsky and M. Black. Black sets questions to Chomsky which are clearly designed to allow Chomsky the opportunity to comment on problems widely felt to be crucial for him. The article on Skinnerianism by R. A. Boakes and M. S. Halliday concentrates mainly upon the tactical aspects of operant research, and is subjected to a vigorous attack by Karl Pribram, presented in the idiosyncratic manner we have learned to expect from him.

The form of this book is such that much of the burden of editorship falls on any reader who hopes that this collection of discrete contributions will yield an intellectual whole. There are few cross references between the basic thematic units, which are tripartite in form (article, criticism and rejoinder). These short interchanges can be of absorbing interest when they concern areas with which the reader is very familiar, but otherwise they can be as much exhausting as educative since (with a few honourable exceptions) there is a dearth of *methodological* formulation. Important questions arise from the debates, but are more often suggested by the particular cross purposes evinced and the immeasurable distance between contributors, than articulated by the writers; and it is often difficult to tell whether criticisms are aimed at specific formulations or at the general position whence these formulations spring. Nearly half the articles are characterised by undefended Popperian assertions.

Methodological questions of general relevance emerge only to be generally neglected. The concept of "an empirical question" is much relied upon but never adequately discussed, yet this is the precise focus of many of the controversies which rack the human sciences. Without any overall methodological analysis the various topics touched on in this book can be called "behavioural" only by courtesy.

These defects are particularly crippling where discussants do not use as evidence each others' putative facts, and without methodological analysis no comparison between these very different case examples is possible. One man's facts can be another man's irrelevancies or superstitions. Thus in answer to Eysenck's view of human personality as a cluster of disparate personality traits, Bannister puts forward a "process" interpretation of action, which Eysenck questions as follows:

"But what does it mean to say that 'the problem remains to be solved in its own terms'? The words are there to be read, but the sentence as such carries no meaning at all".

The conceptual gulf could scarcely be wider or more clearly expressed.

N. H. Freeman

International Symposium on Amphetamine and Related Compounds. Ed. by E. Costa and S. Garattini. New York: Raven Press; Amsterdam: North-Holland. 1970. Pp. xiii + 962. No price.

Amphetamine is a drug that is remarkable for its wide array of powerful and apparently unrelated behavioural effects. It kills appetite, suppresses dreaming sleep, induces psychotic symptoms and is addictive. It has what may or may not be a family of effects on arousal and performance. These in animals range with increasing dose from facilitation of learned performance to locomotor activation and eventually intense motor stereotypies of the foreparts. It also has physiological effects, on the circulation and on fat tissues for example. And of course Speed Kills.

Alpha-MethylPHenylEThylAMINE has the small hook of a methyl group next to its amine group, which makes a nasty spanner to throw in the cerebral works—for amines of ethane attached to ring structures are abundant in certain types of synapse distributed in widely varying numbers through every region of the brain. There can be no doubt that noradrenaline, dopamine and 5-hydroxytryptamine play an important part in some transmission (or, rather, translation) of axonal firing sequences from one neutrone to another. It is now established that amphetamine or some metabolite releases these amines and also probably disrupts their re-uptake into the synapse and breakdown.

Hence the breadth and potency of the drug is not as surprising as the relative discreteness and good remaining organization in the behavioural changes it induces.

This book is a series of 58 succinct and abundantly illustrated papers from a 1969 conference representing all aspects of the metabolism, biochemical pharmacology and psychopharmacology of amphetamine and its chemical cousins. There are lots of data, many of them highly cogent, and clear brief statements of many of the current important hypotheses on amphetamine action. The neurophysiologists are not there, with microelectrophoresis or with recording and tract section. But that may be fair, as the biochemists and chemists and the psychologists and psychiatrists do seem to have made a lot more out of amphetamine. In this area at least, the biochemists often seem happy with behaviour as a major dependent variable—sometimes disguised under the name of "function". However, any such betrothal of biochemistry and behaviour can in the absence of third parties lead to ill-advised intimacies. Theories of central catecholamine action come to be born out of data purely dependent on the final sum of all the effects of a drug over the whole brain. We shall be fortunate indeed if these seem more than infantile monsters once we allow for variation around the brain in the relative importance of the dozen or so known biochemical effects of amphetamine, for the anatomical distribution of behavioural control within the brain, and for primary, secondary, interactive and compensatory consequences of the drug molecule's own effect. Here lies the power of the minority of studies in this book using histochemistry (Fuxe and Ungerstedt), regional neurochemistry (Glowinski), local perfusion (Wise and Stein) and administration directly into brain tissue (Randrup and Munkvad)—interestingly in nearly all cases here in relation to behavioural effects of amphetamine.

There has been considerable success in modifying the amphetamine molecule to augment one effect over the others. This is of great practical importance in the case of anorexigenic properties, and judicious addition of a few carbon and fluorine atoms makes fenfluramine which has these and fat mobilising properties without the stimulant "side-effects" of amphetamine. Then of some theoretical interest are the methoxyamphetamines which have varying hallucinogenic potency, sometimes without appreciable anorexia or stimulation. One wonders how the changed structure achieves the behavioural selectivity—does the action of the drug become restricted to a single amine, to particular brain regions or what?

The amphetamine specialist will have an exhaustive collection of reprints making part of the book redundant, but anyone else wanting a recent detailed survey of much of the area can hardly do better at present than read this book.

D. A. Booth

Physiology of the Retina and Visual Pathway. By G. S. Brindley (Monographs of the Physiological Society, No. 6). 2nd edition. London: Edward Arnold (Publishers) Ltd., 1970. Pp. xi + 315. £4·50.

The first edition of this book was published 10 years ago, and the very considerable progress that has been made in visual physiology since then has meant that the second edition has had to be extensively rewritten. As a result, the length has increased slightly, from 298 to 315 pages, and the price greatly, from £1·50£ to 4·50. The purpose and general plan of the book have, however, remained unchanged. As before, the interest is centred on vision in man, and results from other animals are excluded unless they are considered particularly relevant. Within this area the author states that he has tried to be as comprehensive as possible.

It is, of course, impossible to cover the whole field of visual physiology in detail in a book of this length, and of necessity only a selection of the relevant literature can be reported. Sometimes the selection is very idiosyncratic, and the work of Brindley and his acquaintances receives a disproportionate amount of space. In spite of Brindley's justification of this in the preface, the result can be irritating. To give two instances, he dismisses all human ophthalmoscopic densitometry work done outside Cambridge in eight lines, and in his treatment of the corpus callosum he quotes a little known experiment involving reaction times that was done at Cambridge, but gives little or no mention of the better established experiments and workers in this field (there is, for example, no reference at all to Sperry in this section). The book also contains a number of mistakes. Although some of these may be trivial, such as the statement on p. 2 that Naka and Ruston's work was done on goldfish (it was, in fact, done on tench, rudd, and bream), others are more misleading. For example, on p. 128 it is stated that interocular transfer depends on two memory traces being formed, one in each hemisphere. Brindley reaches this conclusion, which he states categorically, on the basis of an experiment by Butler, but an earlier and more detailed experiment by Myers and Sperry showed that this was not true for difficult taks, although it is probably true for easy ones.

Some mistakes may be inevitable in a book covering a wide field and written by one author, just as a fairly arbitrary selection of facts is unavoidable. These limitations do not prevent the book fulfilling a valuable function. It is very concisely written, and all the topics are dealt with in concrete terms, with little space devoted to speculation. The author has no hesitation in giving his views on the most probable conclusions to be drawn from any set of data. This is refreshing, and often results in a stimulating summary of the topic being dealt with.

W. R. A. MUNTZ

Publications Received

Books

Body Experience in Fantasy and Behavior. By S. Fisher. New York: Appleton-Century-Crofts. 1970. Pp. 690. $13.75.

Visual Perception. By T. N. Cornsweet. New York: Academic Press. 1970. Pp. 475. £7·00.

Intelligence, Creativity and Cognitive Style. By George Shouksmith. London: B. T. Batsford Ltd. 1970. Pp. 240. £2·30.

Foundations of Modern Auditory Theory. Edited by Jerry V. Tobias. (Vol. 1). New York Academic Press Inc. 1970. Pp. 466. £10·50

Biology of Memory. Edited by Karl H. Pribram and Donald E. Broadbent. New York Academic Press. 1970. Pp. 323. £5·35.

The Psychology of Learning and Motivation. Edited by Gordon H. Bower. (Advances in Research and Theory), Vol. 4. New York: Academic Press. 1970. Pp. 339. £6·55.

Learning Theory and Mental Development. By William K. Estes. New York: Academic Press. 1970. Pp. 223. £4·45.

Physiological Correlates of Emotion. Edited by Perry Black. New York: Academic Press. 1970. Pp. 309. £6·30.

Rhythms of Dialogue. By Joseph Jaffe and Stanley Feldstein. New York: Academic Press. 1970. Pp. 156. £3·95.

Norms of Word Association. Edited by Leo Postman and Geoffrey Keppel. New York: Academic Press. 1970. Pp. 467. £11·20.

Visual Pattern Recognition. Peter C. Dodwell. New York: Holt Rinehart Winston Inc. 1970. Pp. 276.

Behavioural Words, the Study of Single Cases. By P. G. Herbst. London: Tavistock Publications, 1970. Pp. 248. £2·50.

CONTENTS

Printed by W Heffer & Sons Ltd Cambridge England

K.S. / AKD
Docket & the
issue to the
Librarian
R Das
27/7/71

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Volume 23 Part 2 May 1971

PUBLISHED FOR THE EXPERIMENTAL PSYCHOLOGY SOCIETY BY ACADEMIC PRESS LONDON AND NEW YORK

NOTES FOR AUTHORS

Only original papers are accepted for publication. All manuscripts should be submitted in *duplicate* to *Professor N. S. Sutherland, Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex, BN1 9QY, England.* One copy must be a *typewritten original* with double spacing and wide margins. Figures should be drawn to professional standards in Indian ink on tracing linen or Bristol Board; the original drawing (plus one copy) should be submitted. Allowance must be made for a reduction in size when the drawing is reproduced (the width of lines as reproduced should be between 0·25 and 0·33 mm). Explanations must be placed in the figure legend; the drawing should contain as little lettering as possible and lettering should be inserted in *blue* pencil or on a tracing paper to fit over the drawing. Tables and figure legends should be typed on separate sheets of paper. Where possible related diagrams should be grouped together to form a single figure.

Intending contributors should pay attention to the *Journal*'s conventions, and in particular to the form of headings, and sub-headings used, and references (in the Reference list titles of journals will in future be printed in full without abbreviation). The *Journal* will use the *Système Internationale* for all measurements (if terms like inches are used, metric equivalents must be quoted). The use of standard abbreviations in the text is permissible (if in doubt, write term in full the first time it is used and give abbreviation in brackets); please note that the *Journal* does not use S for subject.

An abstract must be included. Although abstracts should be as tersely written as possible, they should summarize all the main experimental findings. The length of abstracts will vary with the length and complexity of the experiments reported but should usually be between 50 and 400 words.

An initial decision on a paper will normally be reached within 2 months of receipt. Where only minor revisions are requested, date of publication will be determined by date of receipt of the paper provided a satisfactory revision is received within 1 month of the request for a revision being made. The delay between final acceptance of a paper and its publication will normally be between 3 and 6 months. The date on which the manuscript is received in its final version (apart from minor alterations—see above) is printed.

Contributors receive 25 copies of their paper free. If they require further copies they should apply to the Publishers at the time of returning proofs.

The Quarterly Journal of Experimental Psychology is published for The Experimental Psychology Society by Academic Press, London and New York. Volume 23, 1971 (4 issues): subscription rate £6.50 plus £0.40. postage inland; $16.00 plus $1.00 postage abroad. Subscription orders should be addressed to Academic Press Inc. (London) Limited, Berkeley Square House, Berkeley Square, London W1X 6BA, with the exception of those originating in the U.S.A., Canada, Central America and South America; these should be sent to Academic Press Inc., 111, Fifth Avenue, New York, New York 10003. Application to mail at second-class postage rates at Jamaica, N.Y. 11431.

Undergraduates and graduates of less than three years' standing may obtain the journal at the reduced price of £3.60. They should apply initially to the Hon. Secretary, Dr David Legge, Department of Psychology, City of London Polytechnic, Central House, Whitechapel High Street, London, E1 79F. They should not send any remittance, but enclose suitable evidence of their student status or date of graduation, such as a letter from a university teacher or copy of a graduation list.

Quarterly Journal of Experimental Psychology (1971) 23, 157–167

SHIFTING ATTENTION BETWEEN THE EARS

ANNE M. TREISMAN

Department of Experimental Psychology, Oxford University

Subjects were asked to recall lists of digits presented either alternately to right and left ears or successively to both. Performance was worse with alternating than with successive lists and the decrement was greater with the faster presentation rate. However, there was no interaction between type of presentation and either digit length or list length. The latter finding suggests that the effect arises in perception rather than in memory and is consistent with the idea that it takes time to shift attention or reset the parameters for perceptual processing of alternating monaural speech items.

Introduction

Does it take time to "shift attention" between different sensory inputs? Two experiments in the early 1950's suggested that some finite time was required to alternate from ear to ear in the perception of dichotic auditory stimuli, presented either simultaneously or alternately to the two ears. Broadbent (1954, 1956, 1957*a*, *b*) explored various factors affecting the recall of simultaneous lists of dichotic digits. He found that at fast presentation rates, subjects do better recalling these digits by ear than alternating between the ears to approximate their temporal order. He suggested this was due to two factors: (1) subjects' inability to perceive two digits simultaneously; (2) a difficulty in switching attention rapidly from ear to ear to perceive simultaneous digits serially. Cherry and Taylor (1954) alternated a single speech message between the ears using an electronic switch, and found that intelligibility dropped sharply at about 2·5 Hz switching rate. They suggested this was due to "the mental and the acoustic switching" getting out of phase.

However, later experiments have cast some doubt on the attention shifting interpretation of both these sets of results. For example, Moray (1960), following up Broadbent's experiments, found that subjects were much better able to report dichotic digits when these were presented alternately rather than in simultaneous pairs. Savin (1967) found that subjects still preferred to group successive rather than simultaneous items together when the two tracks of a dichotic tape were presented to the same ear, so that switching between the ears was no longer required. Broadbent and Gregory (1961), in replying to Moray's criticism, agreed that shifting attention between alternating items may be unnecessary when the two ears are the input channels, since these are normally attended to as a single source of information. However, they showed that alternating between the separate modalities of eye and ear still gives poorer recall than grouping items by ear or eye, even when the items are staggered in time rather than simultaneous. Huggins (1964) repeated Cherry and Taylor's experiment, but varied the speech

rate of the recording which he alternated between the ears. He found that as the speech rate increased, so did the rate at which the maximum dip in intelligibility occurred. This suggested that the difficulty was due to interference with a natural perceptual unit of speech (probably the syllable), which had to be received intact on one ear for intelligibility to remain high.

However, other experiments suggest the existence of some finite switching time, for example the limits of successiveness discrimination with stimuli in different sense modalities (Kristofferson, 1967). An analogous difficulty has even been suggested in tasks requiring alternation between different properties of a single stimulus (colour and shape) by Corballis and Phillips (1966), and alternation between different verbal or cognitive categories, such as digits and names of colours, by Broadbent and Gregory (1964). Although shifting of attention would imply very different operations in these different tasks, all of them might require the subject to reset some parameters of his perceptual system, to alter the ways in which he analyses or tests the incoming data. In each case this "resetting" might be expected to take some minimal time, which could, of course, vary with the classes between which attention is shifted (see Treisman, 1969). Differing interference in recall has been suggested as an alternative explanation for the difficulty in cognitive switching (Sanders and Schroots, 1968) but this account appears to ignore the critical importance of presentation rate in producing the difficulty.

The present paper is concerned with shifting between the two ears. It suggests that the idea of a minimum time for alternating between left and right ears may have been dismissed too lightly. There was in fact some evidence of its importance even in Moray's results (1960), since performance was worse (by about 18%) at fast rates with digits presented alternately to the two ears than with digits presented at the same overall rate to both ears at once. Moray did not discuss this difference or test its statistical significance, but it seems difficult to explain on any other basis than by some difficulty in alternating between the two ears. Huggins's suggestion of disruption of syllabic units will not of course apply here, since whole digits were presented to each ear. If Moray's finding is genuine and if we are correct to explain it in terms of some difficulty in switching between the ears, we still need to explain why switching seems easier between alternating dichotic compared to simultaneous dichotic digits. As Moray pointed out, this could be due to the fact that alternating between simultaneous stimuli means switching to a memory trace rather than a present stimulus, and this may lead to poorer performance either because of memory decay, or because switching is more difficult, or because there is some additional interference between simultaneous stimuli either at input or in memory.

With alternating rather than simultaneous stimuli, there seems no *a priori* reason why the two ears should be treated as independent channels rather than a single auditory channel, since they normally function as such when receiving inputs from the same physical source. Selective attention should only be *necessary* when different competing inputs are presented. However, Moray's (1960) results suggest that the ears may function independently when monaurally stimulated even when this is unnecessary to the task. An alternative acount (Yntema and Trask, 1963) would be that subjects divide their attention, but "label" each incoming

item with the ear to which it was presented. This labelling operation is, of course, logically as unnecessary in Moray's task as is switching of selective attention. However, if it did occur automatically and if one adds the hypothesis that perception of order is more difficult between groups with different labels than within groups, this could account for the results. Moray does not state the proportion of errors that were order errors in his alternating compared to his successive conditions, which would clearly be one test of the hypothesis, but another paper makes this a less plausible explanation. Axelrod, Guzy and Diamond (1968) asked subjects to match the perceived rate of monotonic and dichotically alternated clicks and found that the perceived rate was slower for the dichotic presentation, particularly at rates above 7 clicks/sec. They suggested that subjects are unable to follow this rate of alternation and so lose a fixed proportion of the input at high switching rates. In this experiment perception of order is not required.

More recently Moray (1969) has revived the idea of attention-switching to explain the fact that important or relevant signals are occasionally detected in the unattended message in selective listening to continuous messages. He has also pointed out some difficulties in estimating the actual duration required to switch, in view of the absence of evidence on either the "minimum dwell time" on each channel or the phase relations between the switching of the physical stimulus and the switching of the subject's attention. However, it would seem necessary first to establish whether shifting attention does impose any limit on performance before devising experiments to measure the proportion of time allotted to "transit" between or "dwell" within channels.

It therefore seemed worthwhile to replicate the relevant conditions of Moray's experiment and to extend the investigation to a number of other input parameters. The present experiment compared certain combinations of two presentation rates, two list lengths, simultaneous vs. alternating pairs of items, dichotic and binaural presentation and two stimulus durations (using digits which had been compressed by computer). If the limits are due to shifting of attention, presentation rate should clearly be critical and one might expect an interaction between type of presentation (dichotic vs. binaural) and input rate. If memory storage is important in explaining the effect, list length might produce the major interaction with dichotic vs. binaural presentation. Again, if perceptual shifting is involved, lists composed of the shorter compressed digits might allow more time for shifting in the condition with staggered presentation and so show less decrement in dichotic compared to binaural lists.

Method

Stimulus tapes

The stimuli used in all the experimental tapes were recorded at Bell Telephone Laboratories. Each experimental list was generated from a digital master-tape on which one example of each digit at each of two lengths was stored. These items had been pre-recorded by a male speaker at a sampling rate of 10 kc, filtered at 4 kc, compressed or expanded to two standard lengths of 250 and 150 msec and equated in loudness. The method of recording and compressing these items is described in more detail by Sternberg and Pruzansky (in preparation). A stimulus-ordering programme was then used to generate the experimental tapes described below. This programme allowed accurate

synchronization of stimuli on two tracks of an analogue tape-recorder by recording alternate digital samples of the two items to be synchronized. It also allowed the time intervals and rate to be specified very precisely. Intelligibility of the resulting digits in isolation was high: at least 97% for each of the two durations. Each experimental list was preceded by a 250 msec warning tone of 350 Hz, 1 sec before the first item. The digital tapes were converted to analogue form and played back on a two-channel Ferrograph recorder.

The lists used were as follows:

(1) 6 compressed digits (150 msec long) alternating between tracks at a fast rate (with no silence between items, giving an overall rate of 6·7 items/sec).

(2) 6 compressed digits alternating more slowly (with 100 msec silence between items, giving an overall rate of 4 items/sec).

(3) 8 compressed digits alternating fast (at 6·7/sec).

(4) 8 compressed digits alternating slowly (at 4/sec).

(5) 6 "normal" digits (250 msec long) alternating at the slower rate (with no silence between items, giving 4/sec).

(It was, of course, impossible to test the "normal" digits alternating at the fast rate without introducing some temporal overlap between items. Thus this condition had to be omitted.)

(6) 6 compressed digits in simultaneous pairs (one on each track) at a fast rate (with 150 msec silence between pairs, giving an overall rate of 6·7 items/sec).

(7) 6 "normal" digits simultaneous at the fast rate (with 50 msec silence between pairs, giving 6·7/sec).

These particular combinations of the four parameters were chosen because they were the most relevant to the theoretical issues. Twelve different lists of each type were generated.

Procedure

The twelve examples of each of these seven types of list were played in two conditions, once dichotically and once binaurally. The subjects' task was always to recall as many digits as possible in the order in which they were presented. For the simultaneous lists they had to alternate regularly between the ears to report both members of each pair before going on to the next pair. Any effects of right vs. left ears in the dichotic conditions were counterbalanced as follows: with alternating lists the first, third and fifth lists started on one ear and the second, fourth and sixth lists on the other, and subjects recalled the items in the order in which they were presented; similarly with simultaneous lists, subjects were told that the first digit to be recalled from each simultaneous pair was the one on the left (or right) ear in the first, third and fifth lists and the right (or left) ear in the second, fourth and sixth lists. Subjects were asked to say "blank" for any digit they were omitting, so that scoring of correct items in correct positions would be possible. Seven right-handed subjects (volunteer graduate or undergraduate students at Oxford University) were tested; each subject recalled the twelve examples of each type of list in both the dichotic and the binaural conditions. For each type of list, four subjects had 6 dichotic lists, then 12 binaural and another 6 dichotic, and three had 6 binaural, 12 dichotic and 6 binaural. The first list of each group of 6 was not included in the results, so that each subject was scored on ten lists of each type in each condition. Each subject had a different order of list types using a 7×7 Latin square design.

Results

The mean per cent digits correctly recalled in the correct list position in each condition is given in Table I, together with standard deviations. A number of analyses of variance were carried out on the per cent scores after these had been through an arc-sine transformation (normally used on scores which are proportions

of a fixed maximum). In each case dichotic vs. binaural presentation was one factor, trials another and the third was either presentation rate or compressed vs. normal digits. Each factor was tested against its interaction with subjects since all subjects were tested on all conditions. The main results can be summarized as follows.

TABLE I

Mean per cent digits correctly recalled in each condition (standard deviations of subjects' means in brackets)

		Dichotic		Binaural	
		Compressed	"Normal"	Compressed	"Normal"
List length 6					
Alternating lists	—Fast	60 (17·6)	—	85 (14·6)	—
	—Slow	87 (7·5)	85 (10·1)	93 (9·3)	91 (10·5)
Simultaneous lists	—Fast	44 (9·9)	42 (14·0)	53 (5·6)	61 (9·4)
	—Slow	—	—	—	—
List length 8					
Alternating lists	—Fast	41 (14·6)	—	61 (21·9)	—
	—Slow	63 (16·3)	—	77 (18·7)	—

Effects of practice

The main effect of trials did not reach significance in any of the analyses. However on the alternating lists of six digits presented at the slow rate, there was a significant interaction of trials with compressed vs. normal speech ($F = 2{\cdot}93$, $df = 9, 54$ $P < 0{\cdot}01$), and on the alternating lists of six compressed digits, there was a significant interaction of trials with presentation rate ($F = 2{\cdot}18$, $df = 9, 54$, $P < 0{\cdot}05$). In both cases the effect appeared to be due to chance oscillation, with the difficulty of lists in the two conditions happening to be out of phase, rather than any consistent difference in improvement or decrement over trials 1 to 10.

Effects of dichotic vs. binaural presentation, rate of presentation and list length with the compressed alternating lists

Dichotic presentation was significantly worse than binaural both for lists of 6 ($F = 27{\cdot}6$, $df = 1, 6$, $P < 0{\cdot}01$) and for lists of 8 ($F = 10{\cdot}4$, $df = 1, 6$, $P < 0{\cdot}025$). Fast presentation was significantly worse than slow, both for lists of 6 ($F = 28{\cdot}4$, $df = 1, 6$ $P < 0{\cdot}01$) an for lists of 8 ($F = 140{\cdot}1$, $df = 1, 6$, $P < 0{\cdot}001$). The interaction between type of presentation and rate was also significant, but only for lists of 6 ($F = 6{\cdot}6$, $df = 1, 6$, $P < 0{\cdot}05$), and the effect of rate was significant with dichotic but not with binaural presentation, when these conditions were analysed separately. When both list lengths were analysed together, the lists of 8 were significantly worse recalled (in terms of *percentage* of total digits presented) than the lists of 6 ($F = 65{\cdot}9$, $df = 1, 6$, $P < 0{\cdot}001$, and $31{\cdot}0$, $P < 0{\cdot}01$ for dichotic

and binaural lists respectively). List length did not interact significantly with rate of presentation, nor did it interact with dichotic vs. binaural presentation. Comparing the total digits recalled with the two list lengths rather than the percentages, the difference between list lengths disappears almost completely: subjects averaged about 3·5 items correct on the fast dichotic lists, about 5 on the slow dichotic and fast binaural lists and 5·5 to 6 on the slow binaural lists.

The serial position curves shown in Figure 1 are of some interest. The proportions of correct responses in the different positions are probably not independently determined, so that the results of an analysis of variance should be interpreted with caution. However, for what they are worth, they showed a significant interaction of serial position with dichotic vs. binaural presentation ($F = 4{\cdot}7$ $df = 5, 30$ and $5{\cdot}7$, $P < 0{\cdot}01$ and $P < 0{\cdot}001$ for list lengths 6 and 8) and of serial position with the rate of presentation for both list lengths ($F = 4{\cdot}9$ $df = 5, 30$ and $4{\cdot}0$, $P < 0{\cdot}01$, for list lengths 6 and 8, respectively). The effect can be summarized as follows: the serial position curve drops earlier with fast than with slow presentation, and much more steeply with dichotic than with binaural presentation. Thus with fast dichotic lists recall is down to 50 or 60% by position 2 while with slow dichotic, it is still higher than this at position 4. The slow dichotic curves are quite similar to the fast binaural ones, suggesting that dichotic presentation in some way simulates the effect of increased presentation rate (as it would if shifting attention between the ears used some of the time available). The three-way interaction between rate of presentation, dichotic vs. binaural lists and serial position was also significant for both list lengths ($F = 2{\cdot}6$, $df = 5, 30$, $P < 0{\cdot}05$).

The types of error made in the different conditions were also analysed: errors of commission (i.e. substituting a digit not in the list for one in the list) were rare,

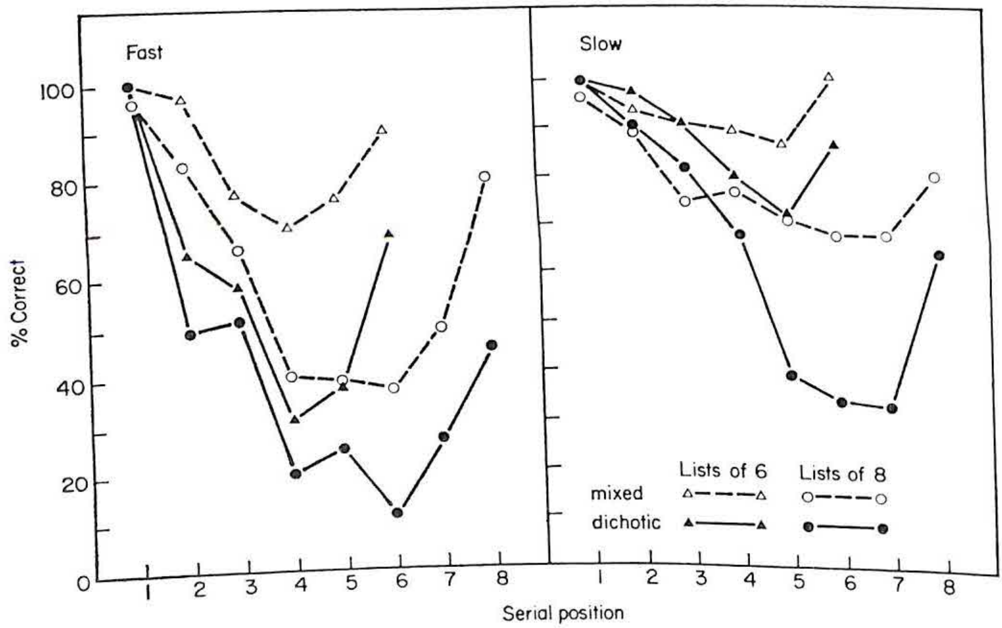

FIGURE 1. Serial position curves for recall of correct items in correct positions with alternating lists of compressed digits. ('Mixed' lists were those presented binaurally.)

averaging 3% in both dichotic and binaural lists. Table II shows the percentages of order errors and omissions in the different conditions (Tables I and II sum to less than 100 in those conditions in which subjects made some errors of commission). In all these rapidly presented lists, recall of order is clearly an important source of difficulty. However, the proportion of total errors which are position (or order) errors is actually lower for dichotic than for binaural lists in the alternating conditions. This is further evidence against the hypothesis that the increased difficulty with dichotic presentation arises in the perception or recall of order rather than in the shifting of attention.

TABLE II

Mean per cent order and omission errors

		Dichotic				Binaural			
		Compressed		"Normal"		Compressed		"Normal"	
		P	O	P	O	P	O	P	O
List length 6									
Alternating	—Fast	22	15	—	—	10	2	—	—
	—Slow	9	4	10	2	6	0	8	1
Simultaneous	—Fast	27	25	28	26	5	38	5	30
	—Slow	—	—	—	—	—	—	—	—
List length 8									
Alternating	—Fast	28	31	—	—	22	13	—	—
	—Slow	15	16	—	—	13	6	—	—

The left-hand figure gives the order errors (items which were presented but recalled in the wrong position) and the right-hand figure the omissions.

Effects of compressed vs. "normal" digits

There was no significant difference between these two types of digits in the mean scores or in the types of error for either the fast simultaneous or the slow alternating lists. There is some suggestion of an interaction with list position in the fast simultaneous lists; performance was usually better on the first of each pair and worse on the second with the compressed digits (78 and 18% compared to 73 and 30% for the normal digits), suggesting that perceiving, or retrieving, the second of each pair was harder with the compressed than with the normal digits. However, this interaction failed to reach significance.

Effects of simultaneous vs. alternating presentation

With the fast lists of compressed digits, simultaneous presentation was worse than alternating ($F = 25{\cdot}6$, $df = 1, 6$, $P < 0{\cdot}01$), and this factor interacted significantly with dichotic vs. binaural lists ($F = 6{\cdot}4$, $df = 1, 6$, $P < 0{\cdot}05$). Several factors could contribute to the difficulty of the simultaneous dichotic lists: (i) alternating between the ears; (ii) central interference between the simultaneous

items; (iii) memory loss in retrieving the trace of the second item if they are analysed serially, and (iv) some difficulty in labelling the ears correctly to alternate between left and right in the required order. If this last factor were the only source of difficulty, the proportion of items correct when one ignores ear-labelling errors should be the same for simultaneous and alternating lists. The present results were therefore re-analysed, ignoring the errors resulting from switches of order within pairs, such as recall in the order of LRRLLR when LRLRLR was required. This analysis showed that switches within pairs were in fact the major cause of decrement with simultaneous dichotic lists: taking the fast dichotic lists of compressed digits, the proportions correct when these switches were ignored were 61% for the simultaneous and 67% for the alternating lists. The apparently superior performance in the binaural compared to dichotic simultaneous lists is thus also due entirely to errors of ear selection in the dichotic case. These errors were of course impossible in the binaural case and much less likely with the dichotic alternating presentation. They averaged 17% in all the dichotic simultaneous conditions and 5% in all the dichotic alternating conditions. The difference between simultaneous and successive lists with *binaural* presentation could also be due to two factors: (i) difficulty in alternating between simultaneous stimuli on a single channel; and (ii) mutual interference or masking between simultaneously presented items. These possibilities need further investigation.

Discussion

The main result confirms that recall of digits presented alternately to the two ears is more difficult than recall of digits presented successively to both. The relative difficulty increases with increased presentation rate, which suggests that the problem arises at input rather than in memory. Rate of recall was not controlled, so there was no time pressure at retrieval. Moreover the relative difficulty of dichotic alternation did not increase with the longer lists, as one might have expected if the source of the difficulty arose in memory storage. This dissociation of the effects due to presentation rate and to increased load on memory supports the perceptual interpretation that there is a limit to the rate at which attention can be shifted between the ears. This attention shifting time would reduce the time available for perception and storage with dichotic alternating presentation, and so simulate the increase in rate of presentation with successive binaural items. An alternative explanation, suggested by Moray (personal communication), might be that attention is divided rather than shifted between the two ears and that this entails some loss in intelligibility. "Peripheral hearing" might show an analogous loss in acuity to that resulting from peripheral vision. This theory would predict that the loss due to dichotic presentation could be simulated by decreasing the intelligibility of the equivalent binaural lists.

Whichever account proves correct, these results do show that separate inputs to the two ears cannot be processed as efficiently as items presented on a single input channel. Why should it be impossible with dichotic stimulation to attend to both ears as if they were a single input channel, as one normally does when presentation is binaural? Presumably these speech stimuli are all analysed by the language systems of the left hemisphere (in right-handed subjects). The difference in

cortical latency at the left hemisphere of left and right ear signals is probably not more than 5 or 10 msec (Simon, 1967), which is unlikely to have much effect other than a slight shift in the "phase" of externally alternating signals, so that at fast rates with no silent intervals there might be a very slight overlap centrally.

With simultaneous stimuli a plausible explanation for the switching limit is Broadbent's suggestion that a selective "filter" must be used to prevent central interference by excluding one stimulus while the other is analysed. Sensory input on one channel may be inhibited or delayed while the other occupies the central speech processing systems. This selective filter then takes time to reset for a new channel. But this account seems less plausible when the stimuli are presented not simultaneously but successively. Even if one argued that at fast rates perception lags behind input and new items must be held up by a filter if their predecessors are still being processed, this would apply as much to successive binaural lists as to alternating dichotic ones.

A more likely explanation, which in its general form would also cover difficulties in dealing with alternating voices or verbal classes, is that certain parameters of the speech recognition system must be reset to cope with words on the left and the right ears. For example if the *form* in which the sensory input arrives were slightly different for the right and left ears, or if it arrived by separate pathways, the tests employed to identify the words would need to be modified for each alternate digit. The fact that subjects were no better at shifting between the compressed than between the "normal" digits is a little surprising on this account; it suggests that either the "resetting" has to wait until the physically different stimulus is present, or that it is delayed until processing of the previous digit is complete and this processing is as slow for the compressed as for the "normal" digits, despite their shorter length. Aaronson (1967) has shown that perception and recall are improved with compressed digits under some circumstances, but this did not occur in the present experiment.

So far the discussion has assumed that incoming alternating stimuli are "perceived" successively and put into a common store, with the difference between dichotic and binaural stimuli arising solely at input. However, when one acts as a subject in these experiments, one has a strong impression, at least with the fast lists, of shifting spatially not only during the presentation but also during retrieval. This impression does not necessarily conflict with the idea that the difficulty is perceptual. As Broadbent (1958) and Aaronson (1967) have pointed out, two strategies are available: (1) to perceive all items as they occur; and (2) if the rate of presentation is too rapid, to store a sensory representation of the input which can subsequently be analysed perceptually. Broadbent has even suggested that both strategies may use the same auditory store. This would imply that perceived items in short term memory are stored not as the information resulting from perceptual analysis, but as the sense data which were analysed and which therefore must be re-analysed either during rehearsal or during recall. (These sensory representations may be identical to the original input or they may be converted into some more abstracted analogue.) If retrieval is from a "sensory tape-recording", it appears in the present experiments to retain its dichotic character at least for the first few seconds. On this hypothesis the distinction between

perception and memory becomes blurred or meaningless, since retrieval involves either perceptual analysis for the first time, or perceptual re-analysis.

The apparent functional independence of the two ears with dichotic stimulation is puzzling in the light of some other recent experiments. Day (1968) showed that when the two ears simultaneously received similar speech stimuli in the same voice, subjects often heard a single "fused" word. For example "poduct" in one ear and "roduct" in the other were heard as "product". Here it seems that these very similar stimuli formed a single binaural input, through fusion of their common phonetic elements. Even more puzzling are some recent findings summarized by Treisman (1970) using simultaneous dichotic stimuli with no shared phonemes at all. When subjects were asked to report one of two accurately synchronized nonsense syllables their responses involved as many switches between the two (for example "TAK" given "TAZ" and "GIK") with dichotic presentation as with binaural presentation, a mean of about 30%. If switching between the ears with alternating stimuli is difficult, as the present experiment suggests, how can it lead to so many spontaneous, apparently unavoidable errors in experiments with simultaneous presentation? Are the ears in some sense less functionally separate when receiving simultaneous than when receiving alternating stimuli? This paradox is as yet unresolved.

This research was supported by a grant from the Medical Research Council. I am also grateful to Bell Telephone Laboratories for the use of their facilities in generating the stimulus tapes.

References

Aaronson, D. (1967). Temporal factors in perception and short term memory. *Psychological Bulletin*, **67**, 130–44.

Axelrod, S., Guzy, L. T. and Diamond, I. T. (1968). Perceived rate of monotic and dichotically alternated clicks. *Journal of the Acoustical Society of America*, **43**, 51–5.

Broadbent, D. E. (1954). Role of auditory localisation and attention in memory span. *Journal of Experimental Psychology*, **47**, 191–6.

Broadbent, D. E. (1956). Successive responses to simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, **8**, 145–52.

Broadbent, D. E. (1957*a*). Immediate memory and simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, **9**, 1–11.

Broadbent, D. E. (1957*b*). Mechanical model for human attention and immediate memory. *Psychological Review*, **64**, 205–15.

Broadbent, D. E. (1958). *Perception and Communication*. London: Pergamon Press.

Broadbent, D. E. and Gregory, M. (1961). On the recall of stimuli presented alternately to two sense-organs. *Quarterly Journal of Experimental Psychology*, **13**, 103–9.

Broadbent, D. E. and Gregory, M. (1964). Stimulus set and response set: the alternation of attention. *Quarterly Journal of Experimental Psychology*, **16**, 309–17.

Cherry, E. C. and Taylor, W. K. (1954). Some further experiments upon the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, **26**, 554–9.

Corballis, M. C. and Philips, R. (1966). Channel by channel report of visually presented "Stroop" items. *Psychonomic Science*, **5**, 465–6.

Day, R. S. (1968). Fusion in dichotic listening, *Unpublished doctoral dissertation*, Stanford University.

Huggins, A. W. F. (1964). Distortion of the temporal pattern of speech: interruption and alternation. *Journal of the Acoustical Society of America*, **36**, 1055–64.

KRISTOFFERSON, A. B. (1967). Attention and psychophysical time. *Acta Psychologica*, **27**, 93–100.

MORAY, N. (1960). Broadbent's filter theory: postulate H and the problem of switching time. *Quarterly Journal of Experimental Psychology*, **12**, 214–20.

MORAY, N. (1969). *Attention: Selective Processes in Vision and Hearing*. London: Hutchinson Educational.

SANDERS, A. F. and SCHROOTS, J. J. F. (1968). Cognitive categories and memory: I. Shifting between categories: II. The effect of temporal versus categorised recall. *Quarterly Journal of Experimental Psychology*, **20**, 370–79.

SAVIN, H. B. (1967). On the successive perception of simultaneous stimuli. *Perception and Psychophysics*, **2**, 479–82.

SIMON, J. R. (1967). Ear preference in a simple reaction time task. *Journal of Experimental Psychology*, **75**, 49–55.

TREISMAN, A. M. (1969). Strategies and models of selective attention, *Psychological Review*, **76**, 282–9.

TREISMAN, A. M. (1970). Perception and recall of simultaneous speech stimuli. In SANDERS, A. F. (Ed.), *Attention and Performance III, Acta Psychologica*, **33**, 132–48.

YNTEMA, D. B. and TRASK, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, **2**, 65–74.

Quarterly Journal of Experimental Psychology (1971) 23, 168–177

STIMULUS AND TASK FACTORS AS DETERMINANTS OF EAR ADVANTAGES

MARK P. HAGGARD AND ALAN M. PARKINSON
The Psychological Laboratory, Cambridge

Two dichotic experiments are reported which dissociate stimulus and task factors in perceptual lateralization. With only trajectories of fundamental frequency as a distinguishing cue, perception of the voicing of stop consonants gives a right ear advantage. Identification of the emotional tone of a sentence of natural speech gives a left ear advantage. If such parameters as fundamental frequency variation or overall naturalness of the speech material determined the direction of an ear advantage, the reverse pattern of results would have been obtained. Hence the task appears more important than the nature of the stimulus.

In an associated paper (Haggard, 1971) we obtained some evidence that *simple* attributes of the stimulus, such as its rate of frequency change, were not relevant to the right ear advantage for speech signals in interaural competition (REA). The present paper examines further the role of simple stimulus factors, and attempts to isolate effects of stimulus and task factors.

Darwin (1969) has extended Kimura's finding (1964) that recognition of musical sequences shows a left ear advantage (LEA) in normal right-handed subjects. Darwin showed, by using pure tone sequences, that the melodic line alone was sufficient to give LEA and that other incidental properties of music are not necessary. He further showed, by inflecting the fundamental frequency on a high quality synthesized syllable ("tea"), that this parameter would give an LEA even in the presence of the overall stimulus conditions normally associated with REA. In this result we cannot be sure whether the stimulus parameter of fundamental frequency is primarily responsible, or the task, which was not a linguistic decision but a drawing out of the pitch contours. The two experiments reported here dissociate these factors.

The first dissociation involves manipulating the fundamental frequency apparently responsible for LEAs and making it cue a linguistic distinction of the type known (generally) to give REA.

A suitable technique involves the rapid trajectories of fundamental frequency (F_0) that can act as acoustical cues to 'voicing' distinctions such as /p/-/b/ or /t/–/d/ (Haggard, Ambler and Callow, 1970). These trajectories are not heard as intonation-like, and are "heard" only in the sense of giving a perception of a voiceless sound (p, t, k) if high and falling or of a voiced sound (b, d, g) if low and rising. It is known that the voicing of stop consonants gives an REA when carried by the normal hierarchy of cues (Shankweiler and Studdert-Kennedy, 1967, 1970). We can then ask if it does so when only the F_0 cue is present. This is the same as asking whether the ear advantage is tied to the most peripheral analyser of all pitch-type information or to the different perceptual processing of linguistic and non-linguistic materials.

The experimental situation shares with experiments yielding LEA an effective dependence of the response upon discrimination of a fundamental frequency pattern. It shares with the experiments yielding REA a response task which is part of a categorical linguistic system rather than a musical system. Hence the direction of any ear advantage will depend upon which of these two factors is the more influential determinant of ear advantages.

Experiment I

This experiment was designed not only to investigate the ear advantage for linguistically significant F_0 trajectories but also the effect of the place of articulation of a consonant upon the perception of its voicing. The latter aspect of the results will be examined elsewhere. The general techniques are the same as those of the previous paper (Haggard, 1971).

Stimuli

The stimuli for the experiment were produced on the Haskins Laboratories' parallel formant synthesizer, and corresponded to four utterances /ba, pa, da, ta/. For a description of some of the acoustical cues in speech see Liberman, Cooper, Shankweiler and Studdert-Kennedy (1967). The place distinction (between the first and second pairs) was created by having different Formant Two and Formant Three (F2 and F3) burst frequencies for the /p, b/ and /t, d/ pairs, and also different onset frequencies for the formant transitions into the vowel. Within these pairs the formant information underlying the place distinction was identical. The 9 msec noise-excited burst was separated from the 36 msec pulse-excited transitions by an 18-msec silent interval.

To create the voicing distinction other physical differences were employed. The most general and effective cue to voicing of a stop consonant is known to be voice onset time (VOT) (Abramson and Lisker, 1967). This is the interval between the burst of a stop consonant and the onset of glottal pulsing in the transition to the following vowel. This VOT was not given different values as the basis of the voicing distinction here because we were concerned with the F_0 trajectory; but it had to have some value throughout the experiment. This was 18 msec, a value that should be ambiguous for the /t/–/d/ distinction although inclined to voicelessness (/p/) in the /p/–/b/ pair. The actual voicing distinction between the members of the pairs was carried by the F_0 trajectory during the same 36 msec as the formants were changing frequency. For sounds intended to be heard as voiced (/b, d/) this was from 72 to 114 Hz, for those intended to be heard as voiceless (/p, t/) from 134 to 114 Hz.

Examples of the four syllables created by this method were made into a 120-trial random-sequence tape with a 6-sec inter-trial interval. Syllables synchronized to within one msec were laid down on the second track of the tape in random sequence apart from the restriction that the syllables were never the same on the two tracks and each syllable was paired equally often with each other one. An extra 6-sec pause delimited sub-blocks of ten trials.

Method

Subjects and procedure

Twelve normal hearing-right-handed Cambridge undergraduates were paid £0.25 per hour to take the test. High quality stereo equipment and Sharpe HA 10 earphones were used. There were 20 practice trials for which results were not disclosed. The 120 trial tape was repeated once and the whole experiment was divided into four 60 trial blocks permitting reversing of earphones and reproduction channels within and across subjects. Subjects were further divided into sub-groups so that three of the six starting with a given reproduction channel to a given ear attended to and reported from that ear on the 1st and 4th blocks, while the other three attended to and reported from that same ear on the 2nd and 3rd blocks. Subjects were informed they would hear examples of four syllables "ba, pa, ta, da". They were told to concentrate hard on the ear specified for each block, reporting only the syllable on that ear, according to the method introduced by Kirstein and Shankweiler (1969). This was done by writing the initial letter onto a numbered response sheet; no blank responses were permitted. This ear-selection method has a great advantage over that of free recall of the stimuli on both ears. It precludes the ear preferences in order of report which, though insufficient to explain ear asymmetries (Bryden, 1967), do occur and complicate both analysis and interpretation.

Results

The marking of each subject's protocol produced two 4 × 4 confusion matrices, one for each ear, each based on 120 trials. The per cent correct scores were obtained for each feature by pooling over pairs of columns and rows differing in the other feature, and Table I gives the means so obtained.

As the place distinction shows the expected REA, there is no doubt about the subjects being potentially able to show REA in these conditions. Although low,

TABLE I

Performance analysed separately for place and voicing

Subject	Place (/p, b/–/t, d/) Right ear	Place (/p, b/–/t, d/) Left ear	Voicing (/p, t/–/b, d/) Right ear	Voicing (/p, t/–/b, d/) Left ear
1	68·8	43·8	52·3	47·5
2	60·0	42·5	56·3	40·0
3	47·5	30·0	55·0	42·5
4	70·0	51·3	55·0	53·8
5	51·3	51·3	55·0	53·8
6	62·5	62·5	46·3	50·0
7	55·0	50·0	57·5	51·3
8	61·3	38·8	55·0	43·8
9	65·0	41·3	61·3	55·0
10	36·3	37·5	57·5	40·0
11	55·0	40·0	52·0	50·0
12	40·0	36·3	55·0	51·3
	Mean REA 13·2 Wilcoxon $T = 1$ $P < 0{\cdot}002$		Mean REA 6·6 Wilcoxon $T = 4{\cdot}5$ $P < 0{\cdot}004$	

Entries are percentage scores.

the overall voicing scores are significantly better than chance ($T = 1$, $N = 12$, $P < 0{\cdot}01$). The low scores appear to reflect overall difficulty of the stimulus situation and the dichotic task rather than any specific inadequacy of the particular cues to voicing. For example, the performance on place is almost as low as on voicing. Haggard *et al.* (1970) obtained 90% correct score using comparable values of the pitch skip cue. Subjects heard the synthetic stimuli as speech-like, and only one questioned their origin. No responses other than the four available were volunteered.

The voicing distinction does show a significant REA. It will be seen that the left ear scores are below chance level. This tendency is significant for the place scores ($T = 6$, $P < 0{\cdot}02$). The only explanation of this is that subjects are unable to obey the intructions to attend to the left ear and fall foul of the fact that the stimuli on the two ears are never the same, consequently tending to have different values on a given feature. These chance and below-chance scores mean that no statements can be made about the magnitude of the REA.

One reason for the low scores on voicing is that the VOT value is only ambiguous for /t/–/d/, tending to countermand the rising pitch cue for /b/ in the /p/–/b/ pair. Scoring for the /t/–/d/ pair only, with correct place of response required, gives a mean percent correct for voicing of 61·7 and an REA of 4·5%. This would be significant for the /t/–/d/ distinction alone but for one subject with a large LEA. His pattern of performance was an interesting reversal of the general pattern; for him the fixed VOT value was more ambiguous for the /p/–/b/ pair than for the /t/–/d/ pair. Hence he performed better on the /p/–/b/ pair, where he also produced a large REA. However for most subjects the REA tendency was greater on the /t/–/d/ pair where the F_0 trajectory cued an otherwise ambiguous voicing distinction. It is not possible without further experiments to say whether this reflects better right ear performance on stimuli not containing contradictory cues or simply greater REA at a higher overall level of performance.

Discussion

The overall right ear advantage obtained for the voicing distinction taken together with previous findings suggests a new generalization: the direction of an ear advantage depends not upon any particular stimulus attribute but upon the total perceptual task. Further stimulus variables would need to be manipulated before we can firmly accept this conclusion, but we should note that here we have employed that stimulus parameter most likely to produce LEA according to previous work.

By varying the stimulus and retaining a task favouring REA, we have retained the REA; it is now necessary to vary the task, retaining a stimulus favouring REA. We thus ask the question, what happens when some more complex speech material, known to give somewhat larger REAs than synchronized synthesized speech (Kimura, 1961) is used, but the task is not a linguistic one, or rather, not one best performed by intervention of linguistic processing. Chaney and Webster (1966) have suggested REA is obtained for isolated vowels sounds when not only identity but inflexion, voice and ear were being reported. However, both Shankweiler and Studdert-Kennedy (1967) and Darwin (1971) did not obtain REA for isolated vowels. It is possible that the Chaney and Webster results stem from insufficient

controls for channel differences in their equipment; earphones were not reversed within and across subjects. Kimura and Folb (1968) have reported REA for backwards speech, but this material is probably still natural enough to be better recognized by partial processing in the usual speech system than by treatment as a non-speech sound. To get round this difficulty we require a simply specified task, preferably a natural one, that is not linguistic; hence the technique of recognition memory for unspecified attributes should be avoided, if any control over the task is to be exercised.

In an attempt to resolve this difficulty, the perception of emotional tone was adopted as the task. A short sentence in any particular emotional tone should be replete with acoustical features for the various linguistic decisions known to yield REA. Hence any LEA for emotions could not be ascribed to cumulative effects of stimulus parameters relevant to the segmental phonemes, and the total task would have to be invoked. On the other hand an REA would tend to argue in favour of stimulus factors as Haggard (1969) reported REA when the task is perception of a contrastive stress in a four-monosyllable sentence. The acoustical variables underlying stress, namely longer-term variations in F_0, intensity and duration are similar to those underlying differences in emotion.

Experiment II

Natural speech (as opposed to synthetic) was used as a stimulus material in this second experiment. This had the advantage of resulting in stimuli that were more natural in general and in the specific sense of including appropriate values of acoustical variables distinguishing the emotions. This would tend to encourage processing of the sentence in a linguistic mode rather than by analytic listening to some simple acoustical attribute, e.g. the intensity alone or the pitch alone. The freedom allowed to the speaker made the length of the stimuli rather variable. This would have resulted in synchronization difficulties had the usual dichotic paradigm been adopted. Therefore, it was decided to present the stimuli to one ear only. A competing stimulus had to be provided on the other ear for two reasons. First, it is necessary to reduce the accuracy with which the stimuli are identified in order to reveal any ear advantage. Second, no ear advantages have been reported without using some form of competing stimulation on the other ear. Experiments of Corsi (1967) suggest, however, that the competing stimulus need not necessarily be of the same class. He obtained REA for digit recall with synchronized white noise as the competing stimulus. Here babble was chosen for this role as it has the same frequency spectrum as speech and some temporal properties that interfere maximally; yet it does not provide a coherent competing message. While the technique of employing an unrelated noise as a competing stimulus has not been investigated sufficiently to show that it gives results identical to the standard technique, we know it does not, for example, guarantee LEA. Its practicability outweighs possible difficulty in interpreting a null result.

Method

Stimuli

The stimuli used consisted of six sentences read in four emotional tones. The sentences used were:

(i) It was last year.
(ii) This is for you.
(iii) Give that to me.
(iv) Put it down there.
(v) Who are those men?
(vi) When can you come?

The emotions used were:

(i) Angry
(ii) Bored
(iii) Happy
(iv) Distressed.

The 24 possible combinations of sentences and emotions were recorded twice each in a random sequence of 48 on a Sony TC 800-A recorder. This recorder has an automatic recording level control which helps to keep the volume of the differnt stimuli similar. The volume compression was not complete, however, and in particular the sentences read in an angry tone were more intense as well as being subjectively louder.

The distracting stimulus used was a babble consisting of many people talking at once. This was recorded on a typical Friday night in a college bar. There also the Sony recorder was used with automatic level control. The babble was dubbed onto the second track of the speech tape and the 48 sentences were spliced out and were joined up with blank tape.

They were made up into an experimental tape consisting of 4 blocks of 48 trials. An interval of 5 sec was left between trials for the subject's response. In addition another tape was prepared giving examples of all the stimuli, and a 48-trial practice tape.

The stimuli were validated by running two subjects under conditions similar to the experimental ones but with no contralateral babble. Any errors should now be due solely to ambiguities in the stimuli. These stimuli produced 100% recognition on sentences and 95% on emotions; thus the task of recognizing the emotion is a reasonable one to present to a subject.

Subjects and procedure

The task of the subject was to indicate by ticking a response sheet both the sentence and the emotion heard. Subjects were instructed to report the sentence first and to leave no blank trials.

After the first and third blocks subjects changed their headphones round. Since the sentences always came over the same channel of the reproduction equipment, this resulted in a change in the ear to which the sentences were presented. Subjects were never in any doubt as to which ear the stimulus would be coming to.

Ten subjects were used, all right-handed normal-hearing undergraduates. They were divided into two equal groups one of which heard the stimuli on the left ear in the first block and the other the right ear. Before the experiment proper was started, the subjects were played the tape containing examples of all sentence-emotion combinations. They were then given 48 practice trials. During these practice trials the relative volume of the sentence and babble channels was adjusted, individually for each subject, until performance level on the emotions was about 50%. Subjects were practised on both ears.

The sound pressure level on the babble channel was about 90 dB at a distance of 1·5 cm from the headpiece. A typical power ratio between the two channels was about 35 dB.

Results

The main results of the experiment are shown in Table II. The entries are the number of correct identifications summed over stimuli for the sentences and the

emotions. The maximum score would be 96 items correct. The "non-contingent" figures are derived from a consideration of all trials of the experiment. The "contingent" figures for the emotions are derived from a consideration of only those trials on which the sentence was correctly reported. One subject (2) obtained 100% correct for sentences.

The mean per cent correct scores on emotions and sentences respectively were 78·3 and 72·3 for the non-contingent figures. In the contingent data accuracy on the emotions was 79·3%.

TABLE II

Summary of results

Subject	Non-contingent data Emotions			sentences			Contingent data emotions		
	L	R	L–R	L	R	L–R	L	R	L–R
1	37	32	5	54	54	0	31	26	5
2	74	71	3	96	96	0	74	71	3
3	73	79	−6	83	75	8	68	64	4
4	79	75	4	84	84	0	73	70	3
5	78	79	−1	75	81	−6	65	68	−3
6	76	72	4	71	73	−2	63	59	4
7	74	64	10	76	72	4	65	59	6
8	76	69	7	72	72	0	61	54	7
9	68	64	4	68	69	−1	53	53	0
10	80	70	10	79	75	4	70	61	9

Entries are no. of correct identifications.

Wilcoxon tests were performed on the non-contingent data for both sentences and emotions. There was no difference between the two ears for recognition of sentences. The emotions gave a left ear advantage of 4·27% ($T = 8$, $N = 10$, $P < 0{\cdot}05$). This did not differ significantly over the types of emotion, although in the case of "angry" there was some tendency for a high score on either ear to override the difference.

The analysis of the contingent data is not straightforward since the number of correctly identified sentences may be different on each ear. Consequently, mere differences between number of emotions correct on each ear cannot be used, although they are shown in Table II for comparison. If we divide the number of emotions correct (given correct sentence) on each ear by the number of sentences correct on that ear then we have a valid basis for comparison. This is done in Table III. The first two columns are the proportions of correct emotions on the right and left ears respectively. The third column is a percentage measure of the difference between these proportions. Since the T-statistic is invariant under monotonic transformation, we can perform the Wilcoxon test directly on these figures. Doing so, we obtain LEA of 4·33% ($T = 4$, $N = 10$, $P < 0{\cdot}02$).

TABLE III

Magnitude of laterality effect in the identification of emotions. Contingent data

Subject	P_L	P_R	$100(P_L - P_R)$
1	0·573	0·481	9·2
2	0·771	0·740	3·1
3	0·819	0·853	−3·4
4	0·868	0·833	3·5
5	0·866	0·840	2·6
6	0·887	0·808	7·9
7	0·855	0·820	3·5
8	0·848	0·750	9·8
9	0·779	0·768	1·1
10	0·887	0·813	7·4

Entries are possibilities of correct response.

Discussion

The LEA for emotions must be interpreted cautiously in view of the failure to obtain any REA for the sentences. This failure could be due to those trials on which the sentence was incorrectly reported being markedly less intelligible with respect to both sentence and emotion. Errors in reporting the sentences could be due to some general acoustic feature of those trials, such as their intensity, unconnected with linguistic significance. Two related features of the data support this interpretation.

First, we saw above that the contingent data attained a higher significance level without yielding a larger LEA. This suggests that the main effect of the contingent analysis is to remove noise from the data, i.e. that the errors on these trials were due to some experimental factor as mentioned above. Second, we may look at the observed probabilities of a correct response to emotion, to sentence, and to both. If the responses to sentence and emotion are independent, then the third of these probabilities should equal the product of the first two. In all cases, bar the subject with 100% correct on sentences, it is greater ($N = 9$, $T = 0$, $P < 0{\cdot}01$). This correlation between accuracy on sentences and emotions implies a fluctuation in audibility of the stimuli or fluctuations of attention.

As far as the ear advantage is concerned, it is not possible or necessary to distinguish between these alternatives at present. In this correlation there is some suggestion that sentence errors are not primarily related to failure of linguistic mode processing, but rather to stimulus or attention factors. This being so, we might not expect REA for the sentences in the present situation. We have made some attempt to obtain linguistic mode processing for the sentences; however, in cases where the sentence was incorrectly reported an above chance score was still obtained on emotions (Mean $= 43{\cdot}5\%$) ($\chi^2 = 80{\cdot}3$, $df = 8$, $P < 0{\cdot}01$). This may indicate an inevitable dissociation of the tasks; it may be psychologically impossible to demonstrate concurrent conflicting ear advantages.

Despite absence of a simultaneous REA for sentences we are left with an LEA for emotional tone. The fact that this was obtained with a natural speech stimulus supports the idea that the perceptual task is of prime importance in determining ear advantages.

While the experimental procedure warrants the simple conclusion drawn, there is one comment about the situation of Experiment II which suggests it may not have been the most sensitive for arriving at that conclusion. The response to the sentence might act to diminish LEA for emotions. Also, if the hemispheric asymmetry is to any extent due to specific competition for an analysing system, a competing stimulus of the same class might have yielded a larger LEA. Or, more generally, the characteristics of a competing stimulus, including its own perceptual lateralization, exert an unknown influence on the magnitude of an obtained asymmetry. Hence the merit of the ear-selection procedure in the present experiment. Future research may be expected to illuminate the effects of the competing stimulus and of the subjects' attentional strategy upon ear asymmetries.

General Conclusion

Taken together the two results give no support for the view that the ear advantage in dichotic experiments is determined by purely acoustical attributes of the stimuli used. In experiments giving both an LEA and an REA the nature of the perceptual task appears to have exerted a greater influence than stimulus attributes. Had the stimulus been the major factor in Experiment I the use of pitch as the cue should have given LEA and in Experiment II the presence of all the cues of a natural utterance should have given REA. The term 'task' should not be interpreted solely as the questions asked of the subject, but rather as the use to which certain stimulus information has to be put. We would not expect response procedure alone to control the ear advantage in the absence of stimulus information relevant to the response required.

We should like to thank Professor O. L. Zangwill for space and facilities, Dr F. S. Cooper of Haskins Laboratories for hospitality and facilities, and JSRU-CESG and Corpus Christi College, Cambridge for research support.

References

BRYDEN, M. P. (1967). An evaluation of some models of laterality effects in dichotic listening. *Acta oto-laryngologica*, **63**, 395–604.

CHANEY, R. E. and WEBSTER, J. C. (1966). Information in certain multi-dimensional sounds. *Journal of the Acoustical Society of America*, **40**, 447–55.

CORSI, P. M. (1967). The effects of contralateral noise upon the perception and immediate recall of monaurally presented verbal material. Unpublished M.A. thesis, McGill University.

DARWIN, C. J. (1969). Auditory perception and cerebral dominance. Ph.D. Thesis, University of Cambridge.

DARWIN, C. J. (1971). Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology*, **23**, 46–62.

HAGGARD, M. P. (1966). Perceptual selectivity in the recognition of speech. Ph.D. Thesis, University of Cambridge.

HAGGARD, M. P. (1969). Perception of semi-vowels and laterals. *Journal of the Acoustical Society of America*, **46**, 115.

HAGGARD, M. P., AMBLER, S. and CALLOW, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, **47**, 613–7.

HAGGARD, M. P. (1971). Encoding and the REA for speech signals. *Quarterly Journal of Experimental Psychology*, **23**, 34–45.

KIMURA, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, **15**, 166–71.

KIMURA, D. (1964). Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology*, **16**, 355–8.

KIMURA, D. and FOLB, S. (1968). Neural processing of backwards-speech sounds. *Science, New York* , **161**, 395–6.

KIRSTEIN, E. and SHANKWEILER, D. P. (1969). Selective listening for dichotically presented consonants and vowels. Paper to E.P.A. Philadelphia, April.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P. and STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431–61.

LISKER, L. and ABRAMSON, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, **10**, 1–28.

SHANKWEILER, D. P. and STUDDERT-KENNEDY, M. (1967). Identification of consonants and vowels presented to the left and right ears. *Quarterly Journal of Experimental Psychology*, **19**, 59–63.

SHANKWEILER, D. P. and STUDDERT-KENEDDY, M. (1970). Hemispheric specialisation for speech perception. *Journal of the Acoustical Society of America*, **48**, 579–3.

Received 8 *October* 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 178–186

CLASSIFICATION WITHOUT IDENTIFICATION IN VISUAL SEARCH

JOAN BRAND†

Institute of Experimental Psychology, Oxford University

Six subjects scanned displays of random consonants for a single target which was (a) another consonant; (b) a given number; or (c) any number. A second group of six subjects took part in three comparable conditions with number displays, and letters or numbers as targets. Scanning time for a number in a letter display or a letter in a number display was more rapid than scanning for a target drawn from the same set as the background. Several unpractised subjects, and all the subjects who practised the task, were able to scan as fast through letters for "any number" as for a specific number, or conversely through digits. The finding of different scanning rates for two precisely physically specified targets, depending on which class they were drawn from, runs counter to an explanation of high-speed scanning in terms of the operation of visual feature analysers. It is suggested that familiar categorization responses may be immediate and may provide the basis for the discrimination of relevant from irrelevant items in rapid visual scanning.

Introduction

In visual search tasks such as those described by Neisser (1963, 1964), the searcher is pushed two ways: he has to scan and reject large numbers of irrelevant items as fast as possible, while attending closely enough to detect a single target when it does occur. He must strike a rather delicate balance between examining the display so carefully that he does not miss targets and "skimming" it rapidly enough so as not to waste time on non-targets. An important factor in the achievement of this balance must be the ratio of non-target to target items in the display; in the experiments cited, this has been very high indeed, usually of the order of 299:1 (Neisser, 1963). With as low a signal frequency as this the subject may well feel he has more to lose by being careful than by being fast; and phenomenal reports at least suggest that his examination of non-target items is cursory indeed. He retains no memory for them, and they even appear during the search as "just a blur" (Neisser, 1964).

A rapid perceptual scanning process which results neither in memory nor even in very detailed percepts, yet which still allows signal detection is plausibly seen as an *abbreviated* form of the full recognition process (Neisser and Beller, 1965) in which only low-level visual feature analysers are activated and more complex levels of analysis held back. It is the aim of this study to see whether high speeds in visual scanning might also be the outcome of a *selective* analysis of the display for more complex stimulus characteristics. Subjects were asked to search for targets which were defined either by an individual description, such as "K", or

† Present address: Medical Research Council, Applied Psychology Unit, Cambridge.

by a class description, such as "any letter". In scanning for an individual target, the subject knows its precise physical appearance, and is in a position to scan the display for physical features of the target by means of rapid parallel feature analysers. In scanning for a class of items, however, the physical features of the target are unknown except as a set of alternatives; and even a comprehensive listing of the set may fail to identify the class. The classes of "letters" and "digits", for example, are made up of much the same simple component features; it wants at least an analysis of combinations of features to distinguish a letter from a digit. Thus when a visual search target is defined only as "any instance" of a particular class of items, such as letters, and not given any physical description, the searcher cannot rely solely on simple feature analysis but needs to interrogate the display at more complex levels. The question this experiment was designed to answer is whether with this type of target definition he can still attain high scanning speeds such as those reported by Neisser.

The class descriptions "a letter" and "a digit" were chosen partly because of the common physical features which make up the two sets; and partly because there is already some evidence that the letter/digit disjunction is an effective basis for the classification of information. Alternate presentation of letters and digits results in poorer short-term retention than grouped presentation (Broadbent and Gregory, 1964; Sanders and Schroots, 1968), and similarly reduces the span of apprehension (Warrington, Kinsbourne and James, 1966). Brown (1960) found that instructions to report only the letters (or digits) from a tachistoscopically presented mixed set were effective if given before but not after presentation; and Posner (1970) found no adverse effect of acoustic similarity on the time subjects took to discriminate letters from digits, in contrast to its effect on the discrimination of one letter from another. These results suggest the possibility that, with these familiar classes, individual items of input may be assigned to a class without detailed processing of the individuals themselves; and that the classifying response may provide the basis of perceptual selection. Such a possibility does not fit easily into models of cognitive operation which assume an hierarchical or sequential order of increasingly sophisticated analysis of inputs, working from the perception of crude physical characteristics to that of semantic content. Two notable examples would be Neisser and Beller's (1965) explanation of their visual search findings, and Treisman's selective attention model (Treisman, 1964; Treisman and Geffen, 1967); within the terms of these models class-membership is clearly a semantic feature.

Method

Twelve subjects took part in the experiment; they were nine men and three women aged between 16 and about 26 years old, members either of Oxford University or of the Oxford College of Further Education. They were paid for their services at the rate of £0.25/hr.

Each stimulus display was a bank of letters or digits arranged in matrix form with 6 characters in every row and 35 in every column. Vowels were excluded from the alphabet of letters and 0 and 1 from that of digits. A PDP8 computer was used to generate and print random sequences of characters in matrix form, and to insert one of a set of possible target characters into one of five predetermined target rows, namely, row 7, 12, 19, 25 or 32. Targets were distributed randomly over the first five spaces in a row. The matrices were

displayed to the subject one at a time at the back of a simple viewing box, 56 cm distant from his eyes as he sat with his head against an eyepiece at the front of the box. What he then saw was a long vertical white field measuring 7 × 17 cm, with an area 2 × 15 cm in the centre occupied by the matrix, and with its top marked by a fluorescent light as a fixation point for the start of the search. At the beginning of each trial the box was dark save for the fixation point, and the subject initiated the trial himself when ready by pressing a key on his left. This illuminated the display by switching on a single 40-W light bulb, out of the subject's field of view at the front of the box, and simultaneously triggered a centisecond timer.

The instructions were to search the display by means of a single steady downward scan, without pausing and without retracing; on finding the target, the subject pressed his right hand key to stop the timer and switch off the light, and then reported the symbol which immediately followed the target. The experimenter checked this for accuracy, recorded the time taken, inserted a new display and told the subject the new target for the next trial. The check on false positives which is provided by the report of the symbol following the target was not used by Neisser in his experiments; a pilot study was run with just two subjects to determine whether this requirement would alter searching strategy in any way, and it was clear that it affected neither slope nor intercept of the search time function. If the subject completed the scan without having found the target he was then allowed unlimited time to locate it; this time was not recorded, nor was the trial repeated, but scored as an omission error.

The experiment had a mixed design; the subjects were randomly assigned to one of two groups of six subjects each. One group always searched through matrices of letters and the other through matrices of digits. Within each group, every subject took part in three conditions, defined by the different types of target:

(1) A physically specified target of the same class as the background (for example, the target "K" in a background of letters).

(2) A physically specified target of the opposite class to the background ("7" in a background of letters).

(3) An unspecified target of the opposite class to the background ("any number" in a background of letters).

The three conditions were given in three blocks, in a different order to each subject in the group, according to the Latin square principle. Within each block, the target appeared in each of the five target rows on five occasions, making 25 trials per block and 75 per session. Subjects were tested individually and were given 20 practice trials at the beginning of each session, 5 for the specific target of the opposite class, 5 for the unspecified target of the opposite class, and 10 for the same-class target.

Results

Search rate

Search rates were calculated as the ratio of the time taken to find the target and its distance from the starting point of the search. Slope values representing search rates in each condition are given in Table I. The linearity of the observed slopes testified to the efficacy of the instructions to search downwards in a single steady scan; none of the subjects claimed to be able to discriminate target rows from non-target rows, but several said that they learned that the target never appeared in the first five or six rows; thus for many of the graphs the intercept may be artificially small.

The data were also subjected to two fourway analyses of variance. The first of these compared the two conditions in which the target was physically specified.

A target drawn from a different class from the background items was more quickly detected than one drawn from the same class (F 1, 10 = 27·9, P <0·001), and there was a significant interaction between class of target and vertical position of target (F 4, 40 = 9·53, P <0·001), indicating a slower *rate* of search for a same-class target. The main effect of the two groups of subjects was not significant, nor were any of its interactions.

The second analysis of variance compared the two types of target, physically specified and unspecified, which differed in class from the background. The significant main effect of specification of target (F 1, 10 = 12·7, P <0·01), and its significant interaction with position of target (F 4, 40 = 9·21, P <0·001) indicated a slower rate of search for an unspecified different-class target than for one which was known prior to the search. Again, neither the main effect of groups nor any of its interactions were significant.

Inspection of the graphs for individual subjects, however, suggested that the second analysis of variance had obscured some interesting differences between individuals. Figure 1 shows the results of three individuals who showed three

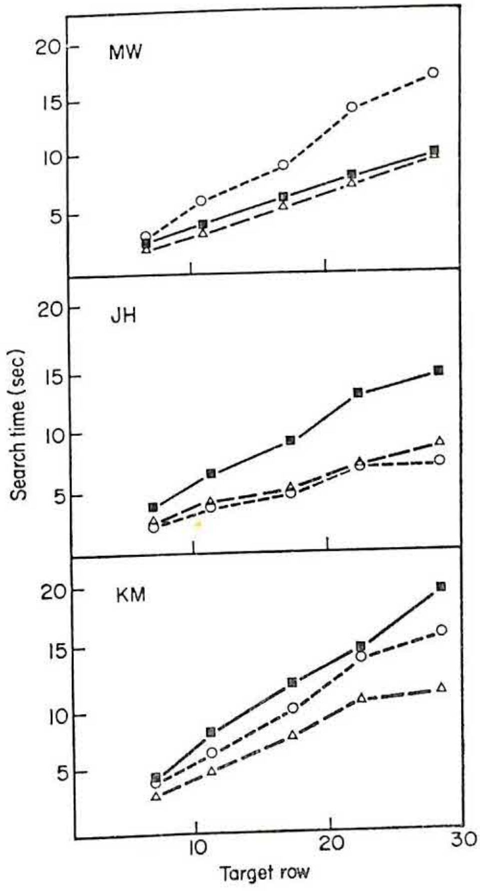

FIGURE 1. Search rates for different targets in letter contexts. Targets: ■–■, letter; △ - - △, known digit; O - - O, unknown digit.

distinct patterns of responding and who seem to represent the whole of the variation over the 12 subjects. For ease of inspection, the three subjects are all from the group who searched though letters; six subjects, three from each group, showed the pattern represented by KM, five (two from the letter group and three from the digit group) that by JH, leaving MW, from the letter group, as an isolated case. MW and JH showed two completely different patterns; for MW the physical specification of the target was the crucial factor; if it was physically specified he scanned for it rather fast and regardless of whether it belonged to the same set as the background or a different one, whereas if he knew only that it was "any number" he was much slower. JH, by contrast, scanned a display of letters faster for a number target than for a letter target, although both were physically specified, and was no faster if he knew exactly which number it was than if he did not. KM seems to fall between these two extremes; unlike MW, he scanned faster for a specific target if it was of the opposite class to the background, but was also helped by specification of the different-class target. In partial confirmation of these results, two separate analyses of variance were carried out on the subgroup of five subjects who showed the pattern represented here by the results of JH. These five subjects searched faster for a different-class specific target than for a same-class one (F 1, 4 = 25·9, P <0·01), but there was no significant difference arising from physical specification of the different-class target (F 1, 4 = 0·77, N.S.).

Errors

The mean percentage omission error rate over all conditions was 6·5%. Table I shows that most of these were made in the conditions where the target was of the same class as the background, and that more were made by subjects searching through digits than by those searching through letters. Analysis of variance

TABLE I

Slope values representing search rates for each of three conditions in each group

	Target conditions			
	Same class	Different class		
		Specific	Unspecified	Error total
Letter Group	0·53 (10)	0·39 (1)	0·50 (7)	18
Digit Group	0·49 (27)	0·30 (6)	0·45 (5)	38
Error total:	37	7	12	56

Figures in brackets indicate the total number of errors made by the six subjects in each condition.

showed that this last difference approached but did not reach significance (F 1, 10 = 4·67, P <0·10), though there was a significant interaction between groups and target conditions (F 2, 20 = 7·86, P <0·01). There was a significant difference between target conditions (F 2, 20 = 21·97, P <0·001), and a Newman-Keuls test

showed that there were significantly more errors in the same-class target conditions than in each of the two different-class conditions, and that these two did not differ.

The effects of practice

It seemed likely that the strategy shown by MW, that of reliance on knowledge of the physical features of the target, might only become feasible on better acquaintance with the particular alphabet and the particular type-face used; Rabbitt (1967) has shown that an important component of the practice effect in visual search is the learning of the critical cues which distinguish a particular target from a particular set of background letters. Four of the 12 subjects, two from each group, took part in four further sessions, making five sessions in all. The results were contrary to this expectation; Table II shows the results from the first and the fifth sessions. MW did not even maintain his initial strategy with practice; by the fifth session his rate of scan for an unknown target of different class had improved more than that for the two physically specified targets, and no longer differed from rate of scan for a specific different-class target. Both were very slightly faster than scan for a same-class target. KM was one of the subjects in the initial

TABLE II

Slope values for the first and fifth sessions of two subjects from each group

			Target conditions		
			Same class	Different class	
				Specific	Unspecified
Digit group	MW	Session 1	0·32	0·33	0·62
		Session 5	0·22	0·13	0·19
	KM	Session 1	0·64	0·38	0·54
		Session 5	0·37	0·13	0·11
Letter group	CM	Session 1	0·33	0·11	0·21
		Session 5	0·19	0·14	0·11
	QS	Session 1	0·39	0·26	0·28
		Session 5	0·28	0·11	0·17

experiment who seemed to combine the two strategies of using physical and class information to scan for the target. By the fifth session, far from showing more reliance on knowledge of the target's physical appearance, he had gone the opposite way—he scanned equally fast for the two different-class targets and more slowly for the same-class target.

CM and QS, the two subjects from the digit group who practised the task, both searched more slowly on the first session for a same-class target and equally fast for the specific and unspecified different-class targets; and both maintained their preference over practice.

Discussion

The results of the experiment may be summarized as follows.

(1) Eleven of the 12 subjects were able to scan faster for a specific number target on a display of letters than for a letter target, or conversely.

(2) Five of the unpractised subjects could scan a display of letters for an unspecified number target as fast as for a specific number target, or conversely.

(3) These differences in search rate were not due to a simple case of speed-accuracy trade-off across conditions; the conditions in which target and background items were both drawn from the same set were also those in which most errors occurred.

(4) The four subjects who practised the task all tended to converge on the strategy of making most use of a distinction in class-membership between target and background, and less of the unique physical representation of the target.

The second and fourth of these results show that subjects were able to scan a visual search display as rapidly for up to 20 physically different targets as for a single one. This finding is not novel in visual search: Neisser, Novick and Lazar (1963) reported scanning for ten different targets as rapidly as for one. The subjects in the present experiment, however, achieved an equivalent rate of scan for the multiple target after little or no practice and without a dramatically large error rate; Neisser, Novick and Lazar's subjects had had fourteen sessions of practice, and made errors at rates of up to 20%. Both the high error rate and the intensive practice required were consistent with Neisser's model of visual search by feature analysis, a low-level, parallel and error-prone operation which depends on thorough learning of features. The lower error rate and little practice of the present subjects might suggest that their performance demands a different explanation; and this is necessitated by result (1), showing different rates of search for two single targets whose physical features are specified.

The multiple physically different targets of the present experiment were of rather a special kind; they were drawn exclusively from a single familiar class of items, they were described to the subject only by means of the name of the class; and the boundaries of the class coincided with the boundaries between the class of target items and the class of background items in the display. Thus a response which *classified* an item as "a letter" or "a digit" without previously identifying it as a particular letter or digit would serve to explain the equivalent rate of search for "any one of 20 letters" or "the letter K". Two of the subjects who practised the task claimed that they were in fact dong this; in scanning through letters, for example, they would set themselves to scan for "a digit" even if told that the target was "3".

This result is in line with Posner's (1970) reaction time data, which indicate that subjects arrive at the classifications "letter" and "digit" without first obtaining the name. It also provides a direct demonstration of the cogency of Rabbitt's (1967) analogy between search and stimulus categorization. The experimental situation differs from Rabbitt's and from Neisser's (1963) in that it makes use of a non-arbitrary, finite and highly familiar category as the basis of discrimination in search, rather than an arbitrary subset of items within a category. The effect of familiarity of the category seems to be that the subject requires little or no practice before being able to use the categorization response as the basis for an efficient and rapid search, with search rate insensitive to the number of items within a class. Indeed, the categorization of letters and digits appears to be so readily available a response that it is not superseded after practice by discrimination based on visual features, although feature learning might be expected to be one of the main effects of practice.

This account of visual search performance resembles Neisser's in that it assumes that the subject does not identify the background material. It differs from Neisser's in that, instead of proposing that identification does not occur because analysis of background items is beyond the level of visual feature analysis, it sees the limitation on the full perceptual process as due to a selective factor; out of the set of possible responses to the presented items the subject focusses on one, namely their membership of a familiar class. The claim that high-speed scanning may rely on a highly-learned classification of input items rather than identification or physical feature analysis cannot be generalized to classifications other than the one used. The digit/letter disjunction is already known to affect cognitive performance in a number of ways, and other classifications, apparently equally familiar, may be less effective. Posner's (1970) finding that acoustic confusability did not affect the discrimination of letters from digits was not replicated for the discrimination of vowels from consonants. Nevertheless, the present results open the possibility of visual scanning by complex characteristics, and emphasize the importance of well-learned classifications in the selection of information.

It should be noted that Neisser and Lazar (1964) obtained a discrepant result; in contexts of letters their subjects were consistently faster in searching for the target "3" than for "any numeral". The present experiment used more single characters than Neisser and Lazar did, and the results in fact seem to vindicate their suggestion that search for "any numeral" was based on a parallel match of all the numbers with display items, and that the speed of the parallel process was held to that of its slowest component, "3" being a distinctive character and therefore one of the faster components.

This work was supported by a Scholarship for Training in Research Methods, No. G 77/1643, from the Medical Research Council. I wish to thank Dr Anne Treisman for her helpful criticisms and discussion, and Dr J. R. Millenson and Geoff Sullivan for writing a computer program for generation of the stimulus displays.

References

BROADBENT, D. E. and GREGORY, M. (1964). Stimulus set and response set: the alternation of attention. *Quarterly Journal of Experimental Psychology*, **16**, 309–17.

BROWN, J. (1960). Evidence for a selective process during the presentation of tachistoscopically presented stimuli. *Journal of Experimental Psychology*, 59, 176–81.

NEISSER, U. (1963). Decision time without reaction time. *American Journal of Psychology*, **76**, 376–85.
NEISSER, U. (1964). Visual search. *Scientific American*, **210**, 94–102.
NEISSER, U. and BELLER, H. K. (1965). Searching though word lists. *British Journal of Psychology*, **56**, 349–58.
NEISSER, U. and LAZAR, R. (1964). Searching for novel targets. *Perceptual Motor Skills*, **19**, 427–32.
NEISSER, U., NOVICK, R. and LAZAR, R. (1963). Searching for ten targets simultaneously. *Perceptual Motor Skills*, **17**, 955–61.
POSNER, M. I. (1970). On the relationship between letter names and superordinate categories. *Quarterly Journal of Experimental Psychology*, **22**, 279–87.
RABBITT, P. M. (1967). Learning to ignore irrelevant information. *American Journal of Psychology*, **80**, 1–13.
SANDERS, A. F. and SCHROOTS, J. J. F. (1968). Cognitive categories and memory span: I : Shifting between categories. *Quarterly Journal of Experimental Psychology*, **20**, 370–2.
TREISMAN, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. *American Journal of Psychology*, **77**, 533–46.
TREISMAN, A. and GEFFEN, G. (1967). Selective attention: perception or response? *Quarterly Journal of Experimental Psychology*, **19**, 1–17.
WARRINGTON, E. J., KINSBOURNE, M. and JAMES, M. (1966). Uncertainty and transitional probability in the span of apprehension. *British Journal of Psychology*, **57**, 7–16.

Received 5 *November*, 1970

Quarterly Journal of Experimental Psychology (1971) 23, 187–196

CHANGES OF PUPIL SIZE AND REHEARSAL STRATEGIES IN A SHORT-TERM MEMORY TASK

DANIEL KAHNEMAN

Hebrew University, Jerusalem, Israel

AND PATRICIA WRIGHT

Medical Research Council, Applied Psychology Unit, Cambridge

Twelve subjects learned lists consisting of 3 groups of 4 items each drawn from vocabularies of digits, colour names or boys' names. There were two conditions of recall, total or partial, and two retention intervals, 3 and 7 sec. A view of the function of rehearsal suggests that rehearsal for total recall should be more intense than for partial recall, but only with a brief retention interval. Measurements of pupillary diameter confirm this prediction. Conditions under which pupillary measurements can serve to test theories of psychological processes are discussed.

Introduction

A salient feature of recent work on short-term memory is the increasing importance which is attached to rehearsal, as a factor to be controlled (Peterson and Peterson, 1959; Waugh and Norman, 1965), as a type of information-processing (Broadbent, 1958; Neisser, 1967; Posner, 1964, 1967; Sperling, 1967), or as an object of empirical study (Corballis, 1969; Mackworth, 1965; Murray, 1967; Ryan, 1969*a*, *b*; Sanders, 1961; Wickelgren, 1964, 1967). In spite of this widespread interest in rehearsal, the denotation of this concept remains elusive, because of the subjective, mentalistic and covert nature of the process. The connotation, however, is well-defined: rehearsal is good, potent and active. In the present paper, we use an objective but highly indirect measure, changes of pupil size, to investigate a situation in which rehearsal may not be good, and in which relative passivity may be a better strategy than intense activity. The study has two objectives: first, to test specific hypotheses concerning rehearsal; second, to illustrate a use of measures of autonomic activity in the study of covert processes.

The suspicion that rehearsal is not always advantageous derives from the idea, which many theorists share, that items spoken by the experimenter and items silently rehearsed by subjects successively occupy a common station on the information-processing line; the P-system for Broadbent (1958), the AIS (Auditory Information Storage) for Sperling (1967), and primary memory for Waugh and Norman (1965). On an interference-theory of rapid forgetting (e.g. Waugh and Norman, 1968) the shared tenancy of short-term storage by items just heard and by items just rehearsed must cause some loss in the retention of the items most recently heard.

This line of argument suggests that the general effect of rehearsal is to improve the recall of early items in the list (e.g. Glanzer and Meinzer, 1967; Posner, 1964; Raffel, 1936), at the cost of some loss in the retention of late, recent items.

The sacrifice of recency for primacy is certainly justified when subjects prepare for serial recall. The first items are crucial in this memory test, because a subject rarely overcomes the confusing effects of an early failure, even when urged to skip over forgotten items and to report anything remembered. When recall is tested by any variant of the probe method, however, the early items are not in any way more important than the most recent. Learning for probed, part recall should therefore be associated with less rehersal activity than learning for total recall, at least when recall is immediate.

If, however, the retention interval is prolonged, subjects can no longer rely on the passive maintenance of primary memory, and they must therefore rehearse, even for partial recall (Anderson, 1960). With a sufficiently long interval, there is no reason to expect any difference in learning strategy between the conditions of partial recall and whole recall, since any retrieval must be from long-term storage. The main hypothesis of this study is that the expected test of recall (whole or part) and the expected retention interval (long or short) interact in their effects on subjects' rehearsal strategy, namely that there is more active rehearsal for whole recall, but only when the retention interval is short.

The measure of rehearsal activity in the present study is an increase of pupil diameter. There is some prior justification for this choice of measure. Kahneman and Beatty (1966) have described the changes of pupil diameter that occur in a short-term memory task; the pupil dilates steadily during the listening phase, and constricts during the report phase, a pattern which presumably corresponds to the course of rehearsal activity. A very different sequence of pupillary responses is observed when nine digits are presented in three groups separated by pauses. Brief dilations now occur only during the pauses (Kahneman, Onuska and Wolman, 1968), in good accord with a view of the effects of grouping on rehearsal by Ryan (1967, 1969*b*). Pupillary diameter is not, of course, a specific measure of rehearsal activity; it is a measure of the intensity of mental effort [for recent references on this point, see Bradshaw (1968), Kahneman and Peavler (1969)]. In the context of serial learning, however, mental effort is most likely to mean rehearsal. We shall return later to the general issues raised by this use of pupillary measurements.

Method

Twelve paid volunteer housewives served as subjects. The lists to be learned consisted of three groups (digits, colour names, boys' names), each of four items. The colour names and the boys' names were selected from vocabularies of ten items each, with which subjects were familiarized during the instruction period. The three types of material were presented in different orders on successive trials.

The material was presented through headphones from one track of a two-channel tape recorder. The same tape was used for all subjects including four experimental conditions: partial recall with a 3-sec interval between the termination of the list and the cue to recall (P-3); partial recall with a 7-sec interval (P-7); and whole recall with the same retention intervals (W-3 and W-7). Each message started with an announcement of the condition. From about ½ sec after the termination of this preliminary instruction, film was taken of the subject's right eye. Tape-recorded pulses from the second channel of the tape recorder

activated the camera so that pictures were taken at intervals of 2/3 sec. This was also the rate at which items were presented. The first item was read to coincide with the third frame, and the last item was read on frame 14. On frame 18 (for P-3) or 24 (for P-7) subjects heard a single word (digits, names or colours) which identified the group that they were to report. The word "repeat" was heard on whole-recall trials (W-3 and W-7). Four additional pictures were taken after the instruction. A total of 28 trials were given. The first eight trials were considered practice, and were not analysed. Subjects were encouraged to report the material in order if they could, but to abandon that attempt if they found it too difficult. The inter-trial interval was not fixed, and subjects were allowed to fully complete their recall before the next trial started. No knowledge of results was given, but subjects often needed, and received, general encouragement.

The subject's left eye was occluded. Pictures of his right eye were taken on Kodak Hi-Speed Infrared 16 mm film with a Siemens instrument camera fitted with a 4-in. lens and a 65 mm extension tube. Lens adjustments were constant for all subjects; the position of the camera was adjusted for each subject to maintain its distance from the corneal bulge at a constant value. Subjects were instructed to look at a fixation mark on the camera, directly over the lens barrel, at a distance of 22 in. from his eye. The developed negative was projected to a flat surface, magnified ×25 and pupil diameter was measured to the nearest mm of projected size.

Results

Pupillary measurements

An average pupillary response curve was computed for each subject under each of the four experimental conditions. The last three trials on which the subject reported at least 6 of 12 items (in the W-3 or W-7 conditions) or 3 of 4 items (in P-3 or P-7) were used. The selection of late trials was justified on the grounds that subjects may require some time to discover the strategies appropriate to the various experimental conditions. The early trials were considered practice for the development of these strategies.

Figure 1 shows the means of the individual curves. A significant preparation effect is apparent on the first picture taken, shortly after the announcement of the experimental condition. Similar effects have been reported previously (Kahneman and Beatty, 1966; Kahneman, Peavler and Onuska, 1968). The interaction that was predicted for rehearsal effort during the trial is apparent as soon as the nature of the trial is announced: the pupil is largest for the W-3 condition (3·97 mm) and smallest for the P-3 condition (3·80 mm). The interaction of recall condition with expected retention interval is highly significant on frame 1 ($F = 20{\cdot}26$, 1/11 *df*, $P < 0{\cdot}01$), but the effect decreases rapidly during the anticipation. Further, as in the other studies which showed a preparation effect on pupil size, this effect vanishes completely once actual work begins. The curves converge within the first 2 sec of the presentation, stay together until the end of the second group of items, and diverge again only during the presentation of the third group. The interaction of recall condition and expected retention interval is significant for pupil size on frame 15 ($F = 5{\cdot}42$, 1/11 *df*, $P < 0{\cdot}05$). The interaction is also significant for the magnitude of the dilation between frame 9 and frame 15 ($F = 5{\cdot}84$, 1/11 *df*, $P < 0{\cdot}05$). The individuals who show the most prominent interaction effect during the presentation of the list are not necessarily those who showed the largest anticipation effect on frame 1. This was established by determining for each subject the magnitude of the difference in pupil size between W-3 and P-3 on frame 1, and subtracting

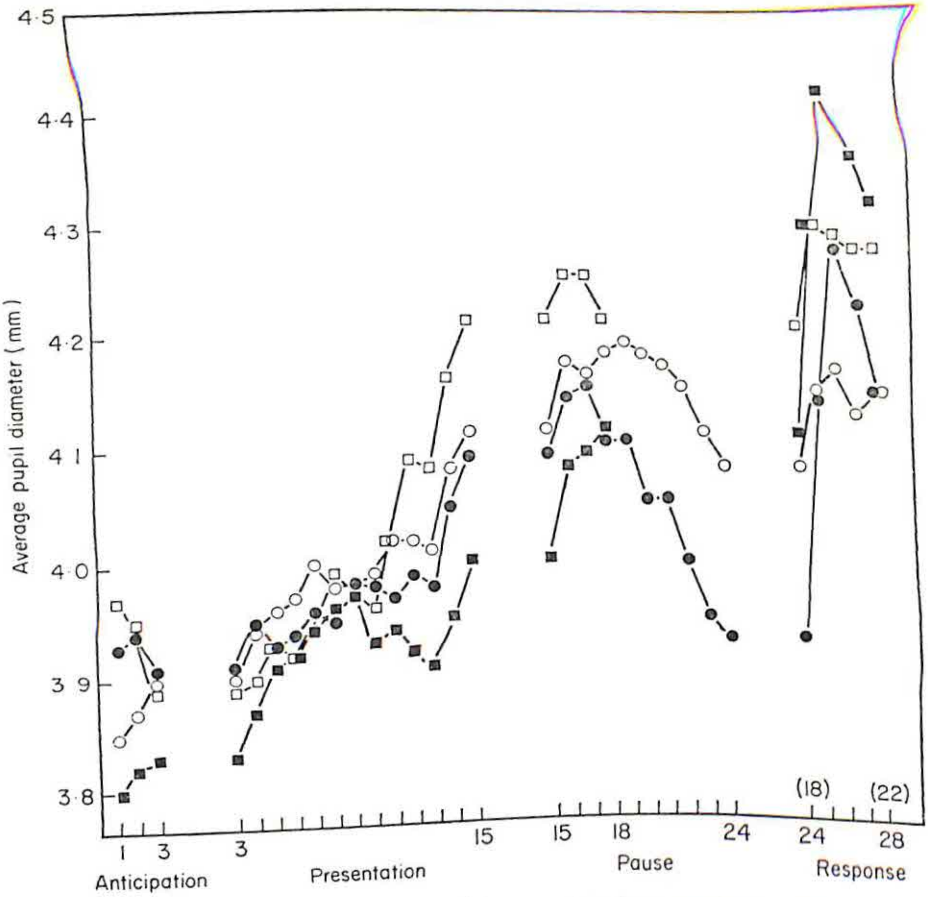

FIGURE 1. Mean pupil diameter for 12 subjects under four conditions: W-7 (○–○), whole recall with 7 sec retention interval; P-7 (●–●), probed recall of a group of 4 items, 7-sec interval; W-3 (□–□) and P-3 (■–■), same conditions with 3-sec retention interval.

from this the difference between W-7 and P-7. A similar difference between the differences was obtained for frame 15. The correlation between these difference scores on frames 1 and 15 was $r = 0{\cdot}03$.

The differences among response curves tend to be maintained during the first few seconds of the retention interval. However, pupil size is significantly larger in the W-7 than in the P-7 condition at the end of the long retention interval, suggesting a relative relaxation in the expectation of partial recall ($t = 3{\cdot}36$, $P < 0{\cdot}01$ for pupil size on frame 24). Finally, the instruction to recall a single group of items causes a much larger dilation on the subsequent two frames than does the instruction "repeat" ($t = 3{\cdot}71$, $P < 0{\cdot}01$, pooling over retention intervals).

Performance measures

Recall data are shown in Figures 2 and 3. Figure 2 shows a primacy effect in whole recall, and recency in partial recall. The effect of recall instruction is highly significant by analysis of variance ($F = 59{\cdot}20$, $1/11$ df, $P < 0{\cdot}001$), as is the interaction of instruction with serial position ($F = 14{\cdot}45$, $2/22$ df, $P < 0{\cdot}001$).

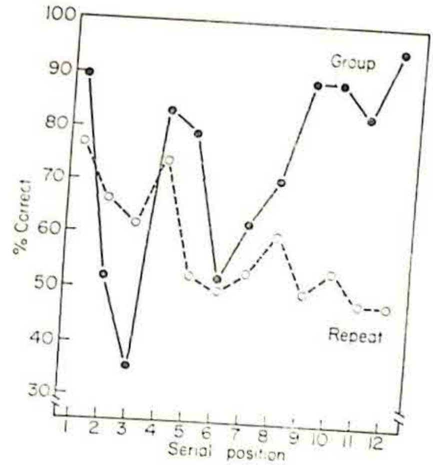

FIGURE 2. Mean percentage recalled in successive groups of 4 items, under four experimental conditions, as in Fig 1.

Retention interval has no significant effect in Figure 2, nor is the interaction of retention interval with serial position significant ($F < 1$). Data for the two retention intervals are pooled, therefore, in Figure 3, which presents a more detailed analysis of position effects.

Figure 3 indicates a marked effect of serial position within each group of items: the first and last items in each group are recalled better than the middle items. This result is highly significant in partial recall ($t = 5{\cdot}30$, $P < 0{\cdot}001$) and marginal in whole recall ($t = 1{\cdot}77$). The magnitude of the serial position effect under the

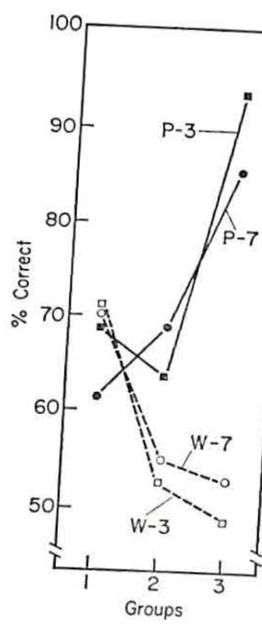

FIGURE 3. Serial position curves for total recall and partial recall of single groups of items. Data for the 3-sec and 7-sec intervals are included.

two conditions can be computed for the first group of items, without confounding of the sequence of presentation with the sequence of report. As Figure 3 suggests the superiority of items 1 and 2 over items 3 and 4 is significantly greater under partial recall than under whole recall instructions ($t = 3{\cdot}59$, $P < 0{\cdot}01$), although the overall number of items recalled is very similar.

Discussion

Theorizing about rehearsal is unavoidably speculative because of the private character of that process. However, enough is known of the characteristics of short-term memory to justify an attempt to determine the conditions under which rehearsal activity may help or hinder recall.

What does rehearsal accomplish? Three effects deserve mention:

(a) Rehearsal of items "recirculates" them through a sub-system of short-term memory, and thereby delays forgetting (Broadbent, 1958; Norman, 1968; Sperling, 1967).

(b) Rehearsal of a short group of items is very effective in reducing the impact of subsequent interpolated activity on recall (Peterson and Peterson, 1959). This is the rationale for the view (Waugh and Norman, 1965) that the main function of rehearsal is to facilitate the entry of items into permanent memory.

(c) Active rehearsal may be viewed as the preparatory fashioning of a response sequence, for a test of memory as for a theatrical performance. It is likely, for example, that the grouping structure which is usually evident in subjects' output of long strings is constructed during rehearsal (Mackworth, 1965; Sanders, 1961; Wickelgren, 1964, 1967). Thus, active rehearsal reduces the effective number of units in store (Miller, 1956) and produces associations between units which are used in subsequent retrieval.

Several conclusions can be derived from this list of functions of rehearsal.

Effects of list-length and retention interval

There is no reason for subjects to rehearse if the quantity of items that must be recalled lies comfortably within the limits of primary memory, provided that immediate recall is required. On the other hand, rehearsal may be needed for a short list if the retention interval is long (for the purposes of recirulation and entry into permanent memory) and for a long list even if recall is immediate (mainly for recirculation and for response organization). There is some evidence in the present study that rehearsal of a list of 12 items is more intense when the retention interval is short (W-3) than when it is long (W-7). Subjects appear to attempt at least one complete rehearsal before they report (Corballis, 1969). They hurry this rehearsal when anticipating an imminent instruction to respond.

Effects of memory task

It is easily seen how rehearsal fulfils the function of preparatory fashioning of a response sequence in learning for serial recall. Since the response in such recall starts at the beginning of the list, it is good strategy to start rehearsing from the

first item, cumulating items in successive rehearsals (Corballis, 1969; Reynolds and Houston, 1964). Naturally, the rate of cumulative rehearsal must increase progressively during the presentation of the list, in order to keep up. The progressive dilation of the pupil during the presentation of a serial list (see Fig. 1) has been interpreted as an indication of this cumulative increase of rehearsal effort (Kahneman and Beatty, 1966; Kahneman, Onusk and Wolman, 1968).

The results of the present study show a cumulative increase of pupil size under conditions of partial serial recall, where the sequence of responses is only partly predictable. This is smaller, however, than in whole recall, when the entire response can be organized during the presentation of the list.

There are learning tasks in which the precise sequence of responses is even less predictable than in the partial recall task; this would be true in both paired associate learning and in learning for free recall. It is perhaps significant that pupillary measurements do not indicate any cumulative increase of efforts in these learning tasks (Kahneman and Peavler, 1969). However, the rate of presentation was much slower in that study than in the present work, and a strict comparison is therefore impossible.

The rule that unpredictability of responses reduces rehearsal activity is not absolute. Broadbent (1957) presented short lists simultaneously to the two ears, and later instructed subjects which list to recall first. He concluded that subjects initially adopted a passive strategy, and refrained from rehearsal because of the unpredictability of the appropriate response. On the second day of testing, however, Broadbent's subjects apparently adopted an active strategy, rehearsed a particular sequence and derived their response, after the cue, from the sequence that they had rehearsed. A long retention interval should encourage subjects to adopt such an active strategy for partial recall (Anderson, 1960). The effects of the long retention interval in the present study confirm this prediction.

Effects of grouping

An exception to cumulative rehearsal in serial learning has been described (Kahneman, Onuska and Wolman, 1968; Ryan, 1969*a*, *b*). When the material to be retained is presented in short bursts separated by pauses, rehearsal appears to be limited to the pauses, and the groups are rehearsed one at a time. This reduction of rehearsal activity is due to the reduced importance of the process of response organization, when grouping in the stimulus sequence provides an adequate blueprint for the response (Mackworth, 1965).

The logic of inferences from pupillary measurements

We have reported an attempt to infer some characteristics of a covert mental process from concomitant autonomic changes. This approach is not widespread, and it requires justification. A detailed discussion of the problem is beyond the scope of this paper, and we can only mention the two central questions that determine the conditions under which pupillary measurements should be considered.

(1) Is there a simpler alternative method? It is quite pointless, for example, to use pupillary measurements to demonstrate that one task is more difficult than

another. Error rates or response latencies provide a more direct and considerably cheaper answer. In the present study, on the other hand, the alternative method was to ask subjects to describe their rehearsal strategies. Our informal attempts to do so yielded a very confusing set of responses; subjects seemed to have no clear idea of the effect that the instructions had on their learning strategies. Pupillary measurements are easiest to justify where introspection is the only alternative.

(2) Can alternative interpretations of the pupillary responses be rejected? Sexual arousal (Hess, Seltzer and Schlien, 1965), problem-solving (Hess and Polt, 1964), the anticipation of having to lift a weight (Nunnally, Knott, Duchnowski and Parker, 1967), surprise (Sokolov, 1963) and anxiety (Hess, 1965) are among the many conditions which elicit pupillary dilations. All these factors must be ruled out if the investigator wishes to use pupillary responses as an indicator of rehearsal activity.

Two types of considerations are relevant to the interpretation of pupillary changes as indications of specific psychological processes. The first is external to the pupillary measurement: the investigator must have a theory of what the subject is doing when a pupillary response is recorded. Thus, we assumed in this paper that subjects' activity during the presentation of the list and during the retention interval can be described as rehearsal. This assumption may be plausible for serial learning. In paired-associate learning, on the other hand, subjects may invest their effort either in rehearsal, or in a search for mediators. The interpretation of pupillary responses must reflect this type of ambiguity whenever it exists.

Internal validating evidence derives from the timing of the pupillary responses during the task. Consider, for example, the alternative interpretation that pupillary dilations only indicate task anxiety. A proponent of that view should try to account for the entire course of pupillary changes from the initial level. Close consideration of Figure 1 will indicate that it is not at all easy to apply the concept of anxiety in this manner.

Some pupillary responses remain, which cannot be interpreted with any confidence. The reader may have noticed our treatment of the pupillary responses which immediately follow the instruction, anticipating the onset of the task. In the absence of any theory for this response (which could, but need not, reflect anxiety), we merely labelled it a preparatory effect.

It is apparent from these considerations that the main advantage of the pupillary measurement technique is that it provides a continuous record of some aspects of subjects' behaviour during the entire course of a trial. Whenever the timing structure of the task is rigidly specified, as it is in most psychological experiments, it becomes possible to time-lock measures of pupil diameter, or of other autonomic indicators (Kahneman, Tursky, Shapiro and Crider, 1969). Each trial then provides information about several aspects of the experimental situation.

References

Anderson, N. S. (1960). Poststimulus cuing in immediate memory. *Journal of Experimental Psychology*, **60**, 216–221.

Bradshaw, J. L. (1968). Pupil size and problem solving. *Quarterly Journal of Experimental Psychology*, **20**, 116–22.

BROADBENT, D. E. (1957). Immediate memory and simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, **9**, 1–11.

BROADBENT, D. E. (1958). *Perception and Communication*. London: Pergamon Press.

GLANZER, M. and MEINZER, A. (1967). The effects of intralist activity on free recall. *Journal of Verbal Learning and Verbal Behavior*, **6**, 928–35.

CORBALLIS, M. C. (1969). Patterns of rehearsal in immediate memory. *British Journal of Psychology*, **60**, 1, 41–49.

HESS, E. H. (1965). Attitude and pupil size. *Scientific American*, **212**, 46–54.

HESS, E. H. and POLT, J. M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science, New York*, **143**, 1190–2.

HESS, E. H., SELTZER, A. L. and SHLIEN, J. M. (1965). Pupil response of hetero and homosexual males to pictures of men and women: a pilot study. *Journal of Abnormal Psychology*, **70**, 165–8.

KAHNEMAN, D. and BEATTY, J. (1966). Pupil diameter and load on memory. *Science, New York*, **154**, 1583–5.

KAHNEMAN, D., ONUSKA, L. and WOLMAN, R. E. (1968). Effects of grouping on the pupillary response in a short-term memory task. *Quarterly Journal of Experimental Psychology*, **20**, 309–11.

KAHNEMAN, D. and PEAVLER, W. S. (1969). Incentive effects and pupillary changes in association learning. *Journal of Experimental Psychology*, **79**, 312–8.

KAHNEMAN, D., PEAVLER, W. S. and ONUSKA, L. (1968). Effects of verbalization and incentive on the pupil response to mental activity. *Canadian Journal of Psychology*, **22**, 186–96.

KAHNEMAN, D., TURSKY, B., SHAPIRO, D. ed CRIDER, A. (1969). Pupillary, heart rate and skin resistance changes during a mental task. *Journal of Experimental Psychology*, **79**, 164–7.

MACKWORTH, J. F. (1965). Presentation rate repetition and organization in auditory short-term memory. *Canadian Journal of Psychology*, **19**, 304–15.

MILLER, G. A. (1956). The magical number seven, plus or minus two. *Psychological Review*, **65**, 81–97.

MURRAY, D. J. (1967). Overt versus covert rehearsal in short-term memory. *Psychonomic Science*, **7**, 363–4.

NEISSER, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

NORMAN, D. A. (1968). *Memory and Attention*. New York: Wiley.

NUNNALLY, J. C., KNOTT, P. D., DUCHNOWSKI, A. and PARKER, R. (1967). Pupillary response as a general measure of activation. *Perception and Psychophysics*, **2**, 149–55.

PETERSON, L. R. and PETERSON, M. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, **58**, 193–8.

POSNER, M. (1964). Rate of presentation and order of recall in immediate memory. *British Journal of Psychology*, **55**, 303–7.

POSNER, M. I. (1967). Short-term memory systems in human information processing. In SANDERS, A. F. (Ed.), *Attention and Performance*. Pp. 267–84. Amsterdam: North Holland.

RAFFEL, G. (1936). Two determinants of the effect of primacy. *American Journal of Psychology*, **48**, 654–7.

REYNOLDS, J. H. and HOUSTON, J. P. (1964). Rehearsal strategies and the primacy effect in serial learning. *Psychonomic Science*, **1**, 279–80.

RYAN, J. F. (1967). *Grouping and short-term memory*. Doctoral dissertation, University of Cambridge.

RYAN, J. (1969*a*). Grouping and short-term memory: different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, **21**, 137–47.

RYAN, J. (1969*b*). Temporal grouping, rehearsal and short-term memory. *Quarterly Journal of Experimental Psychology*, **21**, 148–55.

SANDERS, A. F. (1961). Rehearsal and recall in immediate memory. *Ergonomics*, **4**, 25–34.

another. Error rates or response latencies provide a more direct and considerably cheaper answer. In the present study, on the other hand, the alternative method was to ask subjects to describe their rehearsal strategies. Our informal attempts to do so yielded a very confusing set of responses; subjects seemed to have no clear idea of the effect that the instructions had on their learning strategies. Pupillary measurements are easiest to justify where introspection is the only alternative.

(2) Can alternative interpretations of the pupillary responses be rejected? Sexual arousal (Hess, Seltzer and Schlien, 1965), problem-solving (Hess and Polt, 1964), the anticipation of having to lift a weight (Nunnally, Knott, Duchnowski and Parker, 1967), surprise (Sokolov, 1963) and anxiety (Hess, 1965) are among the many conditions which elicit pupillary dilations. All these factors must be ruled out if the investigator wishes to use pupillary responses as an indicator of rehearsal activity.

Two types of considerations are relevant to the interpretation of pupillary changes as indications of specific psychological processes. The first is external to the pupillary measurement: the investigator must have a theory of what the subject is doing when a pupillary response is recorded. Thus, we assumed in this paper that subjects' activity during the presentation of the list and during the retention interval can be described as rehearsal. This assumption may be plausible for serial learning. In paired-associate learning, on the other hand, subjects may invest their effort either in rehearsal, or in a search for mediators. The interpretation of pupillary responses must reflect this type of ambiguity whenever it exists.

Internal validating evidence derives from the timing of the pupillary responses during the task. Consider, for example, the alternative interpretation that pupillary dilations only indicate task anxiety. A proponent of that view should try to account for the entire course of pupillary changes from the initial level. Close consideration of Figure 1 will indicate that it is not at all easy to apply the concept of anxiety in this manner.

Some pupillary responses remain, which cannot be interpreted with any confidence. The reader may have noticed our treatment of the pupillary responses which immediately follow the instruction, anticipating the onset of the task. In the absence of any theory for this response (which could, but need not, reflect anxiety), we merely labelled it a preparatory effect.

It is apparent from these considerations that the main advantage of the pupillary measurement technique is that it provides a continuous record of some aspects of subjects' behaviour during the entire course of a trial. Whenever the timing structure of the task is rigidly specified, as it is in most psychological experiments, it becomes possible to time-lock measures of pupil diameter, or of other autonomic indicators (Kahneman, Tursky, Shapiro and Crider, 1969). Each trial then provides information about several aspects of the experimental situation.

References

ANDERSON, N. S. (1960). Poststimulus cuing in immediate memory. *Journal of Experimental Psychology*, **60**, 216–221.

BRADSHAW, J. L. (1968). Pupil size and problem solving. *Quarterly Journal of Experimental Psychology*, **20**, 116–22.

Sokolov, Y. N. (1963). *Perception and the Conditioned Reflex*. New York: Macmillan Book Co.

Sperling, G. (1967). Successive approximation to a model for short-term memory. *Acta Psychologica*, **27**, 285–92.

Waugh, N. C. and Norman, D. A. (1965). Primary memory. *Psychology Review*, **72**, 89–104.

Waugh, N. C. and Norman, D. A. (1968). The measure of interference in primary memory. *Journal of Verbal Learning and Verbal Behaviour*, **7**, 617–26.

Wickelgren, W. A. (1964). Size of rehearsal group and short-term memory. *Journal of Experimental Psychology*, **68**, 413–9.

Wickelgren, W. A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, **19**, 97–102.

Quarterly Journal Experimental Psychology (1971) **23**, 197–213

EVIDENCE FOR ALTERNATIVE STRATEGIES OF SENTENCE RETENTION

PATRICIA WRIGHT
Medical Research Council, Applied Psychology Unit, Cambridge
AND DANIEL KAHNEMAN
Hebrew University, Jerusalem, Israel

Measurements of pupil size were taken while subjects listened to sentences and either tried to repeat them (R) or answered a question about them (QuA), after either a 3 or 7 sec retention interval. Pupil dilations were larger for R than for QuA, both towards the end of presentation and during the retention interval. Similar results were obtained by Kahneman and Wright (1971) when comparing total and partial recall of word lists. They attributed the differences in pupil dilations to differences in rehearsal strategy. However, the interaction between recall condition and retention interval was less convincing with sentential material, and an alternative explanation is suggested in terms of the level of abstraction at which the sentences are processed. This interpretation is supported by other evidence relating to the existence of alternative sentence retention strategies.

Pupil dilations failed to reveal phrase juncture phenomena, although two levels of ambient illumination were used in the hope of detecting such effects. In a modified question condition (QuB), where the question was given before the sentence, pupil dilations varied as a function of the part of the sentence providing the answer. These data indicated that people did not begin to frame their answer until they encountered in the sentence those words used in the question.

Introduction

Kahneman and Wright (1971) have suggested that rehearsal varies within a retention task as a function of either total or partial recall being required, and also as a function of the length of the retention interval. Their interpretation is based on the assumption that an increase in mental effort is accompanied by an increase in the diameter of the eye's pupil, and they discuss the validity of this assumption. Specifically, they have shown that during a pause following presentation and before recall, pupil dilations were larger when subjects had to repeat an entire stimulus series than when only part of the series need be recalled; and they further demonstrated that shortening the retention interval increased the magnitude of the pupil dilation duing presentation for total recall but decreased the dilation for partial recall.

Kahneman and Wright were using 12-item word lists, but it is conceivable that an analogous situation exists for experiments concerned with language, particularly studies of sentence retention. Two widely adopted procedures for studying sentence retention are:

(i) requiring verbatim recall of the sentence (e.g. Prentice, 1966);
(ii) requiring an answer to a question about the sentence (e.g. Smith and McMahon, 1970).

It is sometimes implicitly assumed that this latter procedure involves "comprehension" or "understanding" of the sentence (cf. Wright, 1969), and it is conceivable that this technique might involve forms of processing that are essentially different from those required for verbatim sentence retention. It is interesting to note that in the psycholinguistic controversy over kernel sentence forms, many of the studies providing evidence for the existence and importance of kernel strings were studies using verbatim recall (e.g. Mehler, 1963; Savin and Perchonock, 1965); whereas studies raising difficulties for the earliest notions of deep structure frequently used "comprehension" tests (e.g. Gough, 1965, 1966; Slobin, 1966). This observation adds weight to the suggestion that people may have a variety of strategies available for dealing with sentences and that these strategies may differ from each other in certain important characteristics.

The following experiment seeks to reproduce the findings of Kahneman and Wright (1971) pertaining to the difference between total and partial recall, but using sentential material. In addition to the two experimental conditions examined by Kahneman and Wright, a third condition is added. In this third condition subjects need to recall only part of the sentence, but are told before hearing the sentence which part they will be asked for. The purpose of including this third group is to examine the possibility that the mental activity found previously with total recall is the result of some kind of preparation being made for outputting the eventual response, rather than being an optimal rehearsal or retention strategy for that particular task. Output preparation could occur only for total recall in the previous study because only with the repeat instructions did the subject know for certain what his response was to be. But if differing pupil dilations were due to response preparation in this sense, then the third recall condition in the present study should also give large dilations following presentation, even though only partial recall is required.

In view of the demonstration by Kahneman, Onuska and Wolman (1968) that pupil dilations are sensitive to processing of a temporally grouped stimulus input, it was hoped that pupil dilations would be sensitive to the phrase boundaries within sentences. If this were the case, it would mean that pupillometry is a particularly useful technique for psycholinguistic research since it enables statements to be made both about the distribution of mental effort within sentences, and about the relative difficulty of various parts of a sentence without the contaminating effects of recall or other tasks. Therefore the present experiment was repeated at two levels of ambient illumination in order to increase the likelihood of detecting within sentence effects. The lower level of lighting meant that the dilations would be greater (in terms of percentage baseline change) for any given degree of mental effort; but the brighter lighting meant that the pupil constricted more rapidly when not being controlled by mental processing and therefore the cessation of mental effort would be detected more readily.

Method

Through headphones subjects heard a highly modified simple passive sentence; e.g. *The exhausted despondent travellers were frequently cheerfully encouraged by the ambitious intrepid explorer.* After a pause, which was 3 sec for half the subjects and 7 sec for the

others, either the instruction to repeat the sentence was given, or the subject was asked a question such as *Who encouraged the travellers?* 2 sec before each sentence subjects were given one of three sets of instructions. For the three recall conditions the instructions were:

Repeat (R): *Try to repeat this sentence.*
After (QuA): *You will be asked a question about this sentence.*
Before (QuB): *Who encouraged the travellers in this sentence?*

In both the question conditions the question was given in full after the sentence. Subjects were given six practice trials—two with each of the recall instructions and 15 experimental trials, five with each instruction. All 21 sentences differed in content. For each subject the sequence of R, QuA and QuB instructions varied unpredictably from trial to trial, but across subjects each sentence occurred equally often in all recall conditions.

Infrared film of the subject's right eye was taken at the rate of two frames per sec. For details of the apparatus see Kahneman and Wright (1971).

Materials

In order to ensure that pictures were taken at exactly the same points in all sentences the following technique was adopted. A master tape recording was made on a two-channel tape-recorder. On one channel a "master" sentence was recorded, spoken with normal intonation, and on the other channel were a series of beats which were used to trigger the camera. Experimental tapes were then made by listening to the master sentence through headphones while simultaneously speaking the experimental sentence in synchrony with it. A second two-channel tape-recorder recorded the spoken experimental sentence on one channel and simultaneously on the other channel recorded the series of beats from the master tape. By this means all sentences were spoken at identical rates, and pictures were taken in identical places. Ambient illumination was provided by two sources. A 10-in. fluorescent strip was situated approximately 6 in. to the right of and on a level with the subject's right eye. A commercially available red 60 W filament bulb was placed below and slightly in front of the camera lens. The level of illumination was varied by stepping down the voltage from 240 V to 200 V.

The sentential material used in the experiment was designed to slow the input rate as much as possible, and so facilitate monitoring the different parts of the sentence. Consequently all major words were tri-syllabic, and both noun and verb phrases contained two modifiers. Full details of the 15 experimental sentences are given in the Appendix. These constraints upon the material inevitably lessened its "naturalness", and the effect that this may have had is considered below in the Discussion.

Subjects

Forty-eight housewives from the APU subject panel took part in the experiment. They were paid for their services, and were tested in one of four experimental conditions—either high or low illumination, with either a 3 sec or 7 sec retention interval.

Results

The average pupil size under all experimental conditions was computed. Although the peak amplitude of the pupil dilations was greater at the lower level of ambient illumination, there appeared to be no interaction with recall conditions, nor any evidence of phrase boundary effects. Therefore the data from the two brightness conditions were combined for statistical analysis. Figure 1 shows the mean pupil size for the three recall instructions when the retention interval was 7 sec.

Statistical analysis was carried out on a within-subjects basis, by comparing for each person their pupil size under each of the three recall conditions. To obtain

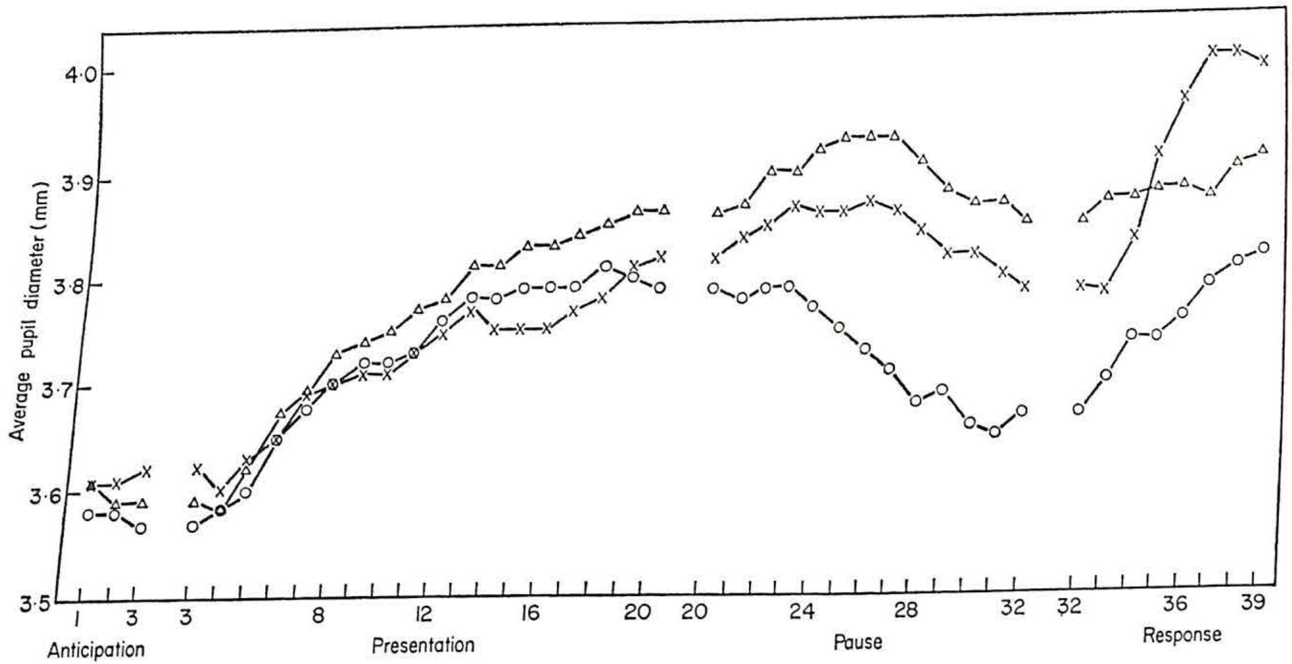

FIGURE 1. Average pupil dilations for the three recall conditions when the retention interval was 7 sec. △–△, Repeat; ×–×, question after; ○–○, question before.

a reliable value for pupil size at any given moment the measurements for that frame, the frame before and the frame after were averaged. The salient points analysed and the results of sign tests at these points are summarized in Table I.

This statistical analysis indicates that there were no significant differences between the three experimental conditions until the middle of the pause. During the pause the dilations with repeat instructions became significantly larger than with either of the question instructions, and this difference continued until it was significantly reversed by the occurrence of the question. Following the question, dilations were largest for QuA.

TABLE I

Results of statistical analysis for data from 7-sec retention interval

	Repeat vs. qu. after	Repeat vs. qu. before	Qu. after vs. qu. before
Start of presentation (frames 3, 4, 5)	N.S.	N.S.	N.S.
End of presentation (frames 18, 19, 20)	N.S.	N.S.	N.S.
Middle of pause (frames 25, 26, 27)	0·01	0·01	0·01 (t test)
End of pause (frames 30, 31, 32)	0·01	0·01	0·01
Instruction to respond (frames 37, 38, 39)	0·01 (*t* test)	0·03	0·01

Similar analysis was carried out of performance with the three recall instructions for those subjects having the 3-sec retention interval. Average pupil sizes are shown in Figure 2, and the results of the statistical analysis carried out as above are given in Table II.

The main points of difference between the results for the two retention intervals are: (1) the earlier appearance of statistically reliable differences between R and QuA with the shorter pause; (2) the much later appearance of significant differences between the two question conditions with the shorter pause.

It will have been noted from a comparison of Figures 1 and 2 that the average pupil size is smaller for people tested with the shorter retention interval. This difference exists prior to presentation of the sentence (frames 1–3) and may be attributable to individual differences between the subjects in the two groups, although Kahneman and Wright offered an explanation of similar differences in terms of anticipatory effects. Perhaps because of this variation in base line pupil size of the two groups, it is difficult to observe the interaction reported in the previous study between retention interval and recall instructions (between R and QuA). From inspection of the graphs, the interaction would appear to be present, since the difference in pupil size at the end of presentation (frame 20) between R

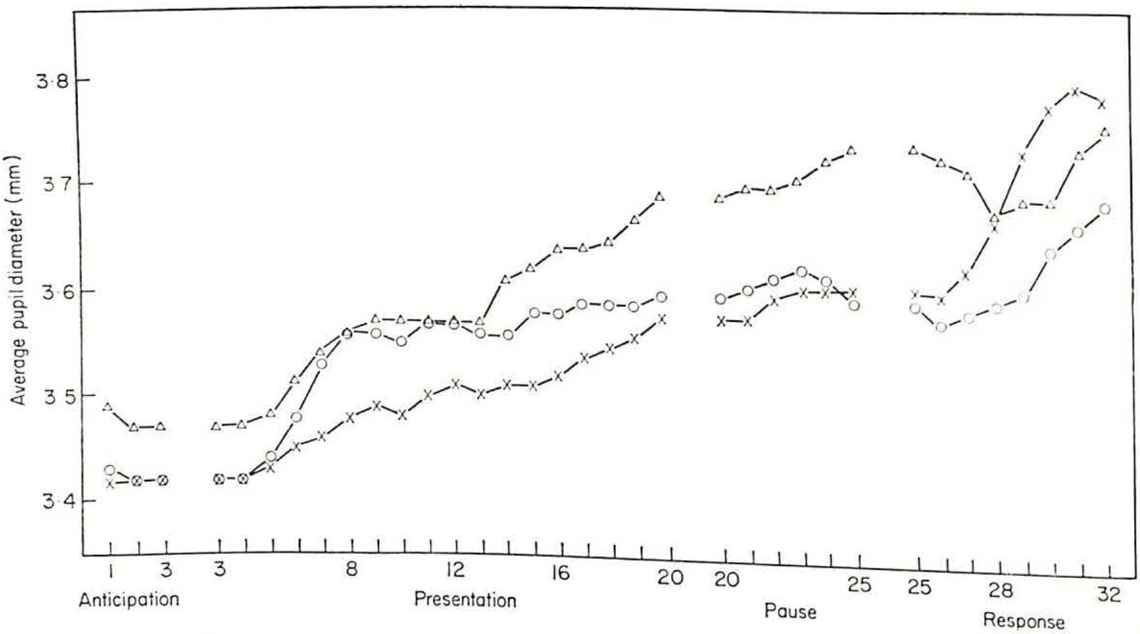

FIGURE 2. Average pupil dilations for the three recall conditions when the retention interval was 3 sec. △–△, Repeat; ×–×, question after; ○–○, question before.

and QuA is larger with the shorter interval than with the longer interval (the differences are 0·12 and 0·04 mm, respectively), but this difference is statistically significant only for the shorter interval. In order to test the interaction directly, the dilation of each subject was computed relative to their own base line pupil diameter; that is to say a measure was obtained of the overall increase in dilation during presentation. However, analysis of variance carried out on these "increase" measures yielded a marginal F ratio for the interaction ($F = 3{\cdot}8$, $1/46$ df, $P < 0{\cdot}05$, one-tailed).

TABLE II

Results of statistical analysis for data from 3-sec retention interval

	Repeat vs. qu. after	Repeat vs. qu. before	Qu. after vs. qu. before
Start of presentation (frames 3, 4, 5)	N.S.	N.S.	N.S.
End of first noun phrase (frames 8, 9, 10)	0·01	N.S.	N.S.
End of presentation (frames 18, 19, 20)	0·01	0·01	N.S.
Middle of pause (frames 23, 24, 25)	0·01	0·01	N.S.
Instruction to respond (frames 31, 32, 33)	N.S.	N.S.	0·01
increase from frame 25 to 31	0·01	N.S.	0·02

Analysis of QuB

Figures 1 and 2 have shown the average pupil size for QuB, but this experimental condition was made up from three different types of question, with the correct answer being either the first or the second noun phrase or the verb phrase. Figure 3 shows the average pupil size in each of these three sub-conditions of QuB with the 7-sec pause, although it must be remembered that there are differences between these sub-conditions both in the words used in the sentence as well as in the part of the sentence being queried, so no unequivocal interpretation of these data is possible.

The points of interest in Figure 3 are the possibilities that there are phrase juncture effects, and that there is a differential difficulty associated with quizzing various parts of the sentence. The first of these points is important because there was no evidence for phrase juncture phenomena in any of the graphs shown in Figures 1 and 2. (The first noun phrase started at frame 4 and ended at frame 8, the verb phrase was from frame 9 to 14, the second noun phrase from just after frame 15 to frame 20.) From Figure 3 it would appear that there are no consistent effects associated with phrase boundaries although the curve for verb phrase questions does show inflections at these points. With reference to the differential difficulty of the various types of question Figure 3 indicates that less mental effort was required when the answer was the second noun phrase than for either of the other questions. This variation in difficulty is also reflected in the error curves shown in Figure 4.

Errors

In Figure 4 it can be seen that for QuB fewest errors were made when the answer was the second noun phrase, and this corresponds to the smaller dilations for this sub-condition shown in Figure 3. However this correspondence does not obtain for the verb phrase and the first noun phrase: the former appearing the more error prone but the latter resulting in larger pupil dilations. It is possible that this lack of correspondence may be partly due to the adjectival or adverbial modifiers within each phrase, because the errors on the nouns and verbs themselves are in the rank order that would be predicted from pupil size (see Fig. 5). These modifiers were often completely forgotten, so it is possible that on some occasions subjects made no attempt to process them. If this were the case, then difficulty would not be reflected in pupil size.

Figure 5 also suggests that people were treating the sentence as consisting of three phrases. Analogous results using lists composed of three taxonomic categories were obtained by Kahneman and Wright, who also obtained the difference in recency effects for the R and QuA conditions shown in Figure 4. Because of the similarity in the distribution of errors for the two pause lengths, a similarity also found in the experiment using word lists, the data from both retention intervals were combined for statistical analysis. A sign test indicated that on the first noun phrase more errors were made in QuA than in R ($P < 0{\cdot}05$), and more errors in R than in QuB ($P < 0{\cdot}05$). On the verb phrase, fewer errors were made with QuB than with either of the other conditions ($P < 0{\cdot}01$) but there was no

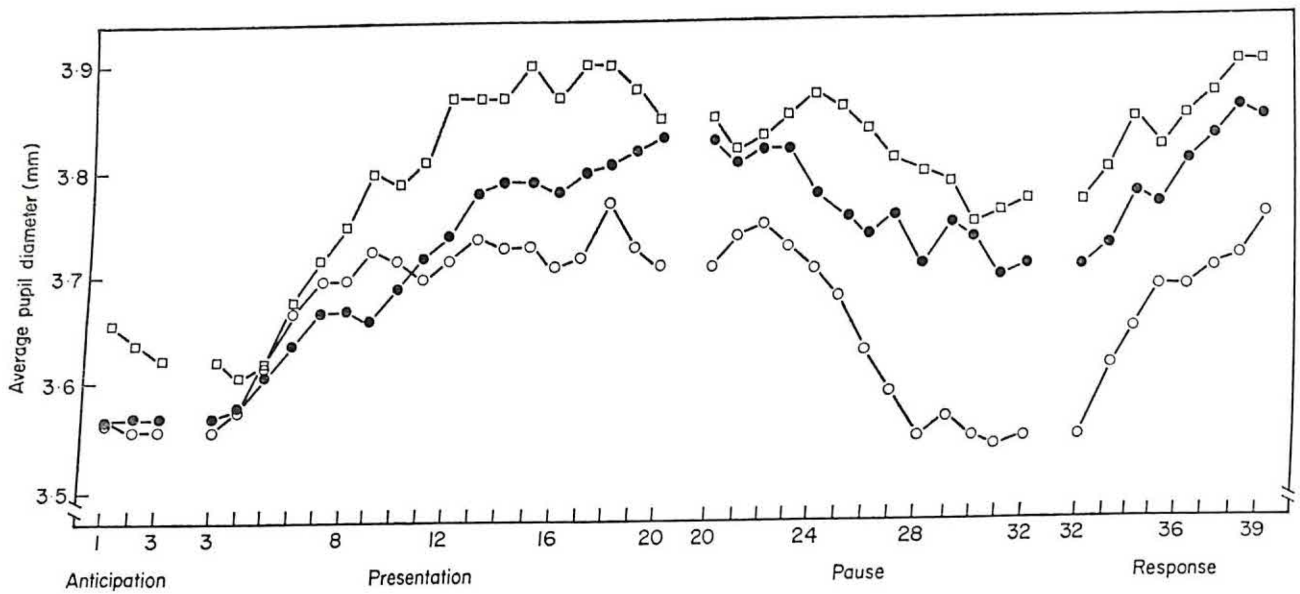

FIGURE 3. Analysis of pupil dilations within question before as a function of the location within the sentence of the answer phrase. □–□, First noun phrase, ●–●, verb phrase; ○–○, second noun phrase.

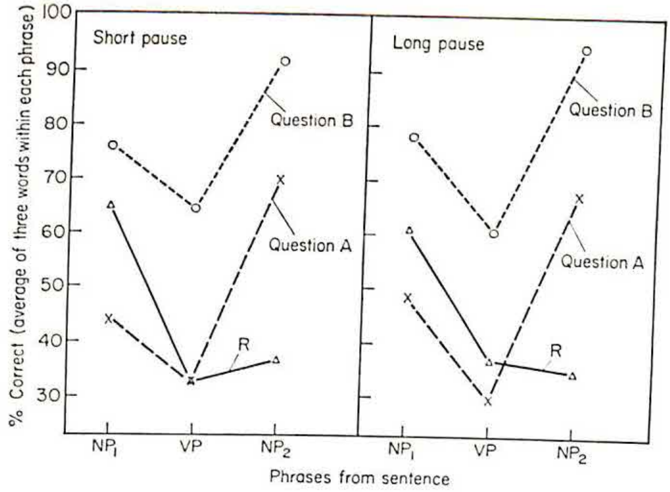

FIGURE 4. Mean per cent errors on each phrase for the three recall conditions, and the two retention intervals.

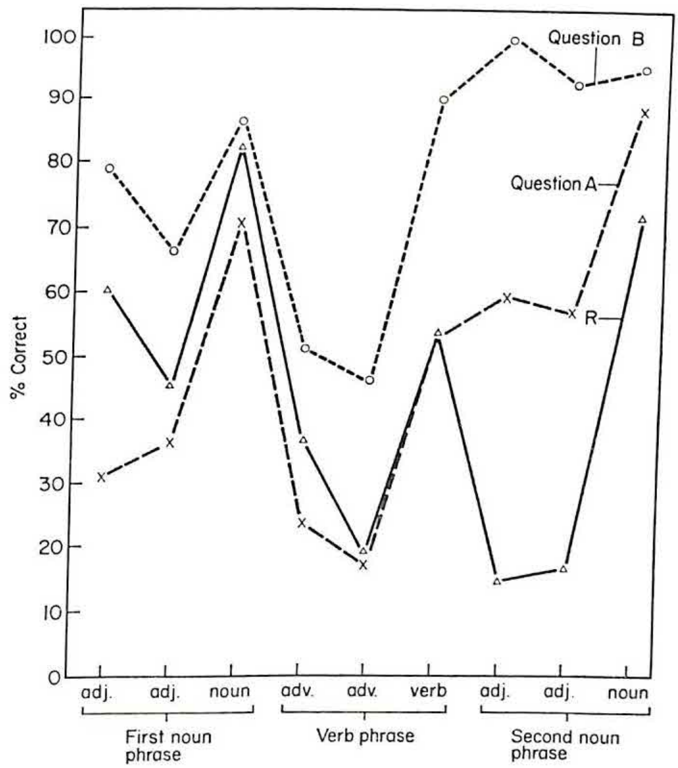

FIGURE 5. Error distribution throughout the sentence for each of the recall conditions, pooling the data from both retention intervals.

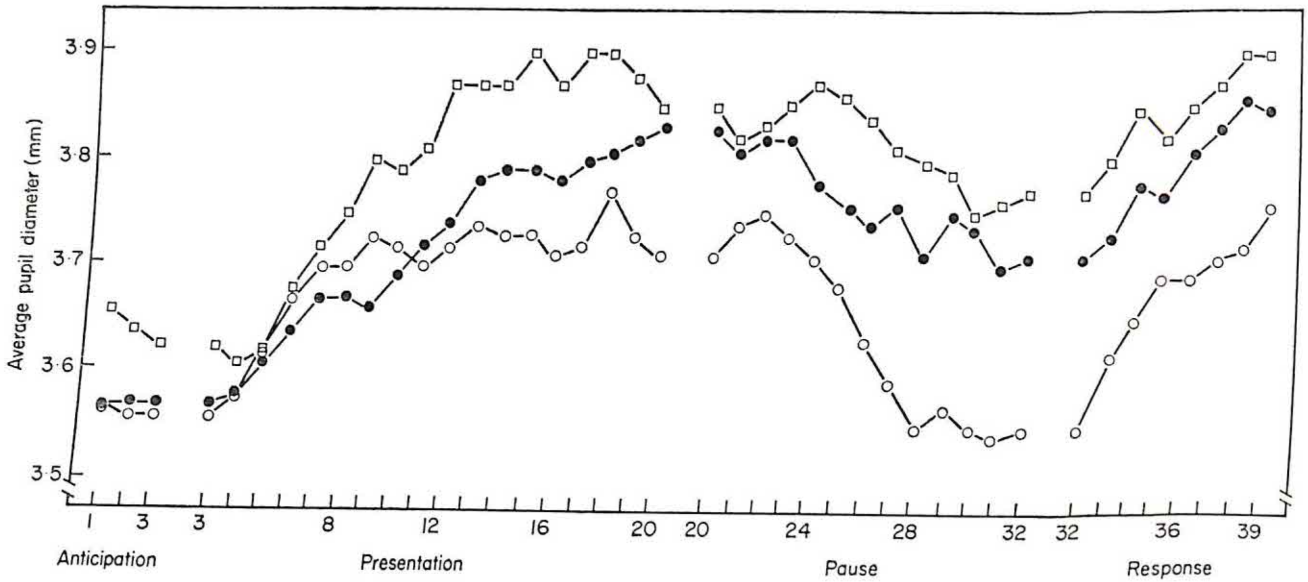

FIGURE 3. Analysis of pupil dilations within question before as a function of the location within the sentence of the answer phrase. □–□, First noun phrase, ●–●, verb phrase; ○–○, second noun phrase.

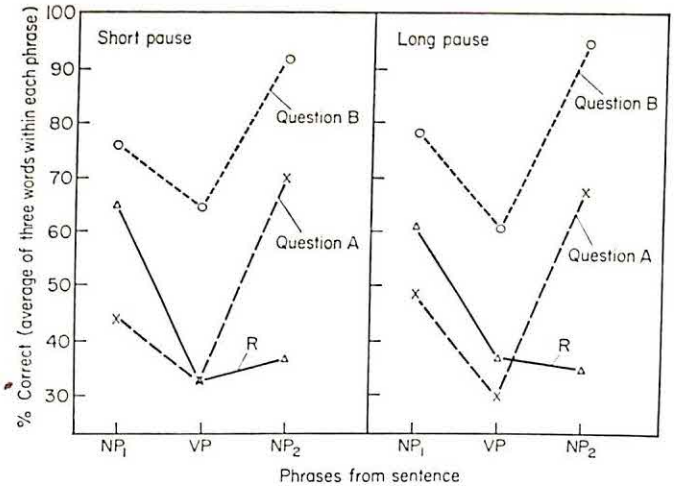

FIGURE 4. Mean per cent errors on each phrase for the three recall conditions, and the two retention intervals.

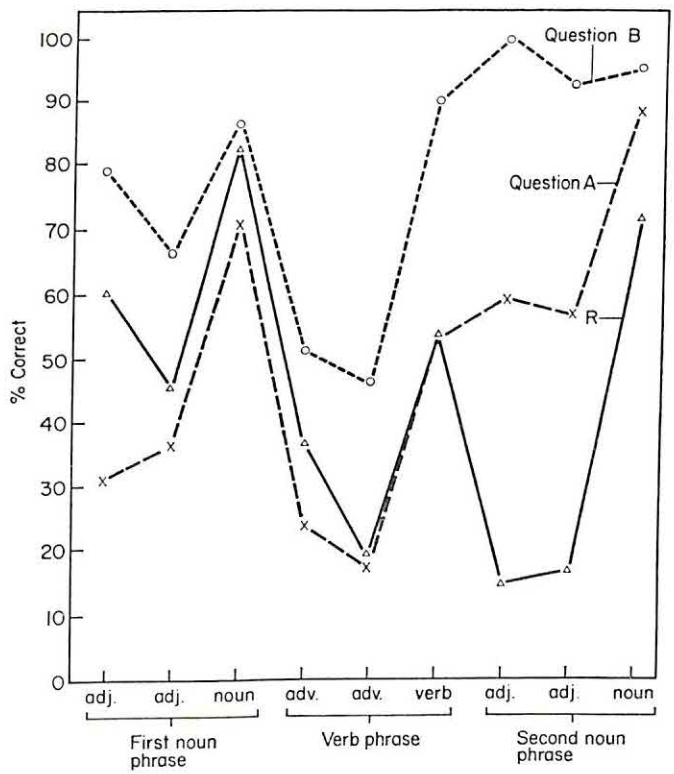

FIGURE 5. Error distribution throughout the sentence for each of the recall conditions, pooling the data from both retention intervals.

statistically reliable difference between R and QuA. On the second noun phrase more errors were made in R than in QuA ($P < 0 \cdot 01$), and more erors in QuA than QuB ($P < 0 \cdot 01$).

As is clear from Figure 4, the shape of the serial position curve is very similar for the two question conditions, but differs from the curve obtained with repeat instructions, suggesting that the pre-and post-questions led to very similar processing by subjects.

Discussion

Sentential material

In many respects the data obtained in the present experiment are similar to the results obtained by Kahneman and Wright (1971) in that they show, for both the long and the short retention intervals, larger pupil dilations in R than in QuA; also the interaction between length of retention interval and magnitude of pupil dilation for different recall instructions is present in so far as the difference between R and QuA is much larger with the 3-sec retention interval. However, the major difference between this experiment and the previous study lies in the nature of the material. It would have been quite plausible to propose that sentences involve kinds of processing that are essentially different from those processes involved in remembering lists of items; or to suggest that the inherent structure of sentences so constrains the way in which they are remembered that other psychological variables would have little effect. But at first sight the present data do not support either of these notions.

It might be possible to argue that the sentences used in the present study were such that subjects processed them as strings of words rather than as natural sentences. The small likelihood of naturally encountering sentences as ponderous as those used here lends force to the argument, and it is not altogether easy to meet this criticism. One can point to the word class effects shown in Figure 5, and suggest that these are characteristic of sentence retention (Martin, Roberts and Collins, 1968); but Kahneman and Wright obtained similar modifications of the serial position curve using unrelated words from three taxonomic categories. Alternatively it might be argued on a priori grounds that answering a question about a sentence necessarily requires that the sentence be processed as something more than a word string. Some support for this assumption is to be found in the data of Figure 4, showing the similarity in the error distributions for the two question conditions. It rather appears that subjects adopted very similar strategies for answering questions in the two question conditions. This point is further supported by the data shown in Figure 3, where dilations are larger towards the end of presentation when the answer is the first noun phrase rather than when it is the second noun phrase. It seems as though people waited until they encountered the words which were used in the question before they started framing their answer. This delay probably accounts for the higher error rate when the answer is the first noun phrase. For questions about the verb phrase also, it seems that subjects waited until the words which were used in the question occurred in the sentence. All the sentences were passive and all verb questions mentioned the actor; e.g. *What did the explorer do?* Consequently the "repeated" words for

verb questions occurred in the second noun phrase of the sentence, just as they did for questions about the first noun phrase. This probably accounts for the similarity in pupil dilation during presentation for these two question conditions. However, it was pointed out in the Results section that the data shown in Figure 3 are difficult to interpret because of the confounding of content differences between the sentences and differences in the location of the answer phrase. Therefore these data can only be considered tentative support for the existence of sentence processing.

A more elegant demonstration that the sentences were not simply being treated as word strings would have been to include evidence of syntactic effects such as the relationship between voice of sentence and question reported by Wright (1969). Unfortunately this was not done in the present experiment because varying the grammatical from of sentences inevitably affects the time it takes to say them (Glucksberg and Danks, 1969); and such variation would have complicated both the experimental design and the analysis of the pupil data. It will be recalled from the section on procedure that great pains were taken to ensure the similarity of intonation pattern and pausing across different experimental sentences.

However if one is prepared to accept that the materials used in the present study are validly termed sentences, then, from the varying distribution of mental effort in R and QuA, shown in Figures 1 and 2, the conclusion seems necessary that people do not approach the problem of remembering a sentence in only one way. And it is these different strategies of retention which are reflected in the different distributions of errors obtained.

Processes underlying retention strategies

In the study using word lists Kahneman and Wright suggested that the observed differences in pupil size for total and partial recall tasks were predominantly due to variations in rehearsal activity and the advantages of adopting a passive rehearsal strategy when only partial recall is required. The distribution of errors in the present experiment is so similar to the earlier data that this explanation might apply here as well, but alternative interpretations are also possible. Kahneman and Wright pointed out that, strictly speaking, pupil dilations were reflecting the intensity of mental effort. Equating this mental effort with rehearsal was theoretically defensible for serial learning tasks. With other kinds of task, theoretical considerations might lead to different interpretations of changes in pupil size. For sentence retention, it seems tempting to suggest that perhaps people were working predominantly with the surface structure in R, but with either the deep structure or with semantics in QuA and QuB. If this were to be the case then clearly the similarity between the present findings and those obtained using word lists must be dismissed as simply fortuitous.

The error distributions for QuA and R could be taken as support for this alternative interpretation. Deep structure processing would tend to emphasize the logical subject in the sentence which in this case, since the sentences were passive, would give the recency effect observed with the questions. On the other hand, surface structure processing would emphasize the surface subject and so would give the greater primacy effect with the obtained R instructions. However,

any adequate explanation must also account for the overall larger pupil dilations with the R instructions. At first perhaps it does not appear intuitively reasonable that surface structure processing, which involves little more than assigning phrase boundaries within the sentence, should require greater "mental effort" than the seemingly more abstract processing involved in determining deep structure relations. But Bever, Lackner and Kirk (1969) have shown grounds exist for thinking that surface structure is in fact abstracted and assigned later than the appreciation of the logical relations among the content words in the sentence. Moreover, Bever (1970) has reported evidence of a tendency for surface structure characteristics to become more apparent as the retention interval increases, although the time interval he considered was of the order of days rather than seconds. Examination of the effect of pause length in the present study indicated that the differences in pupil size between R and QuA during presentation were much smaller and statistically non-significant for the long retention interval. This could be taken as support for the suggestion that test condition R leads to a dominance of surface structure processing, as does lengthening the retention interval. Thus it is arguable that the different strategies being used by subjects in the present study relate not to rehearsal in the senses used by Kahneman and Wright (i.e. recirculation of items, or response organization) but to the levels of abstraction at which the sentences are being processed on different occasions.

Determinants of retention strategies

The major concern of this paper is not to examine alternative interpretations of the above data in any detail. With our present knowledge of the psychology of memory and of language, it seems unlikely that a single unequivocal interpretation is possible. Rather, the following disussion explores some implications of this notion of variable retention strategies, particularly those implications which have a bearing on the experimental investigation of language.

In the introduction, reference was made to the controversy over whether sentences are remembered in terms of abstract properties such as the kernel strings proposed by early transformational grammar, or in some other way, and to the conflict of data which has arisen when different experimental procedures have been adopted. It was suggested that there might have been basic differences between techniques requiring retention and those demanding comprehension. From the present data it seems probable that even within retention tasks a range of alternatives exists.

Although unparsimonious, this suggestion is seen to be feasible in the light of other studies of sentence retention. There are a number of studies which are most easily interpreted on the assumption that subjects were working with the surface structure of the sentence. (For example studies using verbatim recall—Johnson, 1965; Prentice, 1966; studies using probe latency techniques—Suci, Ammon and Gamlin, 1967; Ammon 1968; Kennedy and Wilkes, 1969; studies using sentence-picture material—Gough 1965, 1966; Turner and Rommetveit, 1968.) But there are a number of other studies which show that subjects have taken acount of the less obvious grammatical relations within the sentence (e.g. the logical subject/object relation—Blumenthal, 1967; Blumenthal and Boakes,

1967; similarly the subject/object relations within nominalizations—Rohrman, 1968; Rohrman and Polzella, 1968; and the relations between non-adjacent words —Weisberg, 1969). Even for single words outside a sentence context, grammatical derivation is influential (Briem and Lowenthal, 1968). And there are yet other studies suggesting that subjects may sometimes concentrate predominantly on the meaning of the sentence to the extent that details of the syntactic form and the actual words used in a sentence are easily forgotten (Sachs, 1967; Fillenbaum, 1966; Begg and Paivio, 1969).

But these processing differences do not operate in isolation. There are studies which stress the interaction of syntactic and semantic effects (for passives—Slobin, 1966 and Herriot, 1969; for negatives—Wason, 1965; Wales and Grieve, 1969). Moreover there is evidence indicating that people do not have to make an all or nothing choice about the level of processing they adopt, but that they may process a sentence in more than one way following a single presentation. For example, Clark and Clark (1968) collected data about the retention of sentences such as *Before he swiped the cabbages he tooted the horn.* Clark argued that his data gave evidence for the retention of word order information (i.e. surface structure), knowledge about the temporal order of events (i.e. semantic information) and the main subordinate relation between the clauses (i.e. syntactic information). In passing it should be noted that multiple processing does not necessarily mean parallel processing. Bever (1970) has shown that there appear to be changes in the way sentences are processed as a function of the delay prior to recall even though the length of the retention interval was not known at the time of presentation. This might be evidence for the existence of some serial processing of sentences. For further discussion of some possible variants of "multiple processing" models of sentence processing see Smith and McMahon (1970).

The alternatives just considered were strategies differing primarily in the level of linguistic analysis being undertaken. However the picture becomes yet further complicated since the range of alternatives must also include strategies specific to the psychological processes of STM. There are a number of such strategies. This is evident both from the pupil data of Kahneman and Wright, and the results of Crowder (1969) who showed that the serial position curve for the retention of nine-item lists changes when subjects know prior to presentation how many items there will be in the list. From this finding Crowder inferred that prior knowledge had changed the strategy subjects adopted for remembering the list. For sentences, a similar finding has been reported by Wright (1968). The presence of syntactic effects was found to vary with the range of syntactic forms presented during the experimental session. Therefore, returning to the problem of accounting for the present data, it is clearly difficult to distinguish whether the differences in pupil size are predominantly due to the psychology of memory or to the psychology of language. At the very least, attention must in future be paid to the dynamic interaction of these two systems.

Very little is known about the conditions which influence or determine the sentence retention strategies which may be adopted in different experimental situations. Nor is it known which of these strategies most closely relate to the forms of sentence processing used outside the laboratory. Without doubt, before

an adequate model can be developed for any performance in which memory is involved, it will be necessary to have a better understanding of the options open to people, a more detailed knowledge of how many strategies for retention might be available. There are few answers to these questions, for they have seldom been asked.

Usefulness of pupil measurements

Besides indicating that different levels of mental effort are involved in different sentence retention tasks, it was also hoped that pupil dilations in the present study would provide information about within sentence processing such as might occur at phrase boundaries. However, there was no evidence of any phrase juncture effects, which seems surprising in view of the likelihood that auditorily presented passive sentences are processed as three major "chunks" corresponding to the two noun phrases and the verb phrase (Wright, 1970). Moreover the present data contrasts with the findings of Kahneman, Onuska and Wolman (1968) for performance with a temporarily grouped digit list, where intermittent growth in pupil size was clearly apparent. Perhaps the absence of "scalloping" in the present data is indirect evidence that the sentence processing which subjects were carrying out in this experiment differed from the kind of processing undertaken with non-sentential material.

Within the range of ambient illumination studied in the present experiment the sensitivity of the pupilometry technique did not appear affected by variations in base line pupil diameter. Even with the brighter lighting the treatment effects were sufficiently large to be clearly apparent, and the dimmer lighting was not dark enough to create problems with ceiling effects. This should not be taken to imply that individual differences were always within these limits. But, since a within-subject design was used, idiosyncratic variations due to over-constricted or over-dilated pupils were not critical to the analysis. The effects of somewhat lower levels of ambient illumination have been explored by Bradshaw (1969), who examined pupillary responses in reaction time tasks.

A discussion of the usefulness of pupilometry as a psychological technique was undertaken by Kahneman and Wright (1971). The only additional feature that it seems necessary to point out here is that for experimental investigations of language, pupil size is clearly a less contaminating dependent variable (contaminating in the sense of artifactually encouraging particular processing strategies) than the more conventional error data obtained from retention tests or the latency data obtained from true/false quizzes. Perhaps as more sophisticated apparatus becomes available (e.g. Mackworth, 1968) it may be possible to make a constructive link between studies of eye movements and the distribution of mental effort in tasks such as reading, skimming, proof checking, etc.

Certainly when one considers the variety of tasks in which there is some element of language processing, it seems intuitively reasonable to expect that there might be important differences in the strategies that people adopt under different circumstances. Perhaps it seems regrettable that the use of fairly atypical linguistic material means that very little can be said, on the basis of this experiment, about the psychological processes which might predominate when people are dealing with

natural language. But the recall and question techniques used in this study are probably just as artificial as the sentences.

Typically it is "applied" psychologists who worry about the general validity of experimental findings (e.g. Chapanis, 1967); but it may be time that some of us who are interested in the psychology of language shifted the focus of our research from examining what people *can* do, in the extremity of a laboratory environment, to discovering what people *do* do, unselfconsciously when processing language. It is possible that the recent emphasis on studying underlying linguistic competence has sometimes blurred the difference between these two quite separate questions.

References

AMMON, P. R. (1968). The perception of grammatical relations in sentences: a methodological explanation. *Journal of Verbal Learning and Verbal Behavior*, **7**, 869–75.

BEGG, I. and PAIVIO, A. (1969). Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior*, **8**, 821–7.

BEVER, T. G., LACKNER, J. and KIRK, R. (1969). The underlying structure sentence is the primary unit of speech perception. *Perception and Psychophysics*, **5**, 225–34.

BEVER, T. G. (1970). The integrated study of language behaviour. In MORTON, J. (Ed.), *Biological and Social Factors in Psycholinguistics*, Logos Press.

BLUMENTHAL, A. L. (1967). Prompted recall of sentences. *Journal of Verbal Learning and Verbal Behavior*, **6**, 203–6.

BLUMENTHAL, A. L. and BOAKES, R. (1967). Prompted recall of sentences. *Journal of Verbal Learning and Verbal Behavior*, **6**, 674–6.

BRADSHAW, J. L. (1969). Background light intensity and the pupillary response in a reaction time task. *Psychonomic Science*, **14**, 271–2.

BRIEM, J. and LOEWENTHAL, K. (1968). Immediate recall of nominalisations and adjectivalisations. *Psychonomic Science*, **11**, 209.

CHAPANIS, A. (1967). The relevance of laboratory studies to practical situations. *Ergonomics*, **10**, 557–77.

CLARK, H. H. and CLARK, E. V. (1968). Semantic distinctions and memory for complex sentences. *Quarterly Journal of Experimental Psychology*, **20**, 129–38.

CROWDER, R. G. (1969). Behavioural strategies in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, **8**, 518–23.

FILLENBAUM, S. (1966). Memory for gist: some relevant variables. *Language and Speech*, **9** 217–27.

GLUCKSBERG, S. and DANKS, J. H. (1969). Grammatical structure and recall: a function of the space in immediate memory or of recall delay? *Perception and Psychophysics*, **6**, 113–7.

GOUGH, P. B. (1965). Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior*, **4**, 107–11.

GOUGH, P. B. (1966). The verification of sentences: effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior*, **5**, 492–6.

HERRIOT, P. (1969). The comprehension of active and passive sentences as a function of pragmatic expectations. *Journal of Verbal Learning and Verbal Behavior*, **8**, 166–9.

JOHNSON, N. F. (1965). Linguistic models and functional units of language behavior. In ROSENBERG, S. (Ed.), *Directions in Psycholinguistics*. New York: Macmillan Co.

KAHNEMAN, D., ONUSKA, L. and WOLMAN, R. E. (1968). Effects of grouping on the pupillary response in a short-term memory task. *Quarterly Journal of Experimental Psychology*, **20**, 309–11.

KAHNEMAN, D. and WRIGHT, P. (1971). Changes of pupil size and rehearsal strategies in a short-term memory task. *Quarterly Journal of Experimental Psychology*, **23**, 187–196.

KENNEDY, R. A. and WILKES, A. L. (1969). Analysis of storage and retrieval processes in memorizing simple sentences. *Journal of Experimental Psychology*, **82**, 396–8.

MACKWORTH, N. H. (1968). The wide-angle reflection eye camera for visual choice and pupil size. *Perception and Psychophysics*, **3**, 32–4.

MARTIN, E., ROBERTS, K. H. and COLLINS, A. M. (1968). Short-term memory for sentences. *Journal of Verbal Learning and Verbal Behavior*, **7**, 560–6.

MEHLER, J. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, **2**, 346–51.

PRENTICE, J. L. (1966). Response strength of single words as an influence in sentence behaviour. *Journal of Verbal Learning and Verbal Behavior*, **5**, 429–33.

ROHRMAN, N. L. (1968). The role of syntactic structure in the recall of English nominalizations. *Journal of Verbal Learning and Verbal Behavior*, **7**, 904–12.

ROHRMAN, N. L. and POLZELLA, D. J. (1968). Recall of subject nominalizations. *Psychonomic Science*, **12**, 376.

SACHS, J. S. (1967). Recognition of semantic aspects of discourse. *Perception and Psychophysics*, **2**, 437–42.

SAVIN, H. B. and PERCHONOCK, E. (1965). Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, **4**, 348–53.

SUCI, G. J., AMMON, P. and GAMLIN, P. (1967). The validity of the probe-latency technique for assessing structure in language. *Language and Speech*, **10**, 69–80.

SLOBIN, D. I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, **5**, 219–27.

SMITH, K. H. and MCMAHON, L. E. (1970). Understanding order information in sentences: some recent work at Bell Laboratories. In D'ARCAIS, G. and LEVELT, W. (Eds.), *Advances in Psycholinguistics*, Amsterdam: North Holland Co.

TURNER, E. A. and ROMMETVEIT, R. (1968). Focus of attention in the recall of active and passive sentences. *Journal of Verbal Learning and Verbal Behavior*, **7**, 543–8.

WALES, R. J. and GRIEVE, R. (1969). What is so difficult about negation? *Perception and Psychophysics*, **6**, 327–33.

WASON, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, **4**, 7–11.

WEISBERG, R. W. (1969). Sentence processing assessed through intrasentence word associations. *Journal of Experimental Psychology*, **82**, 332–8.

WRIGHT, P. (1968). Sentence retention and transformation theory. *Quarterly Journal of Experimental Psychology*, **20**, 265–72.

WRIGHT, P. (1969). Transformations and the understanding of sentences. *Langugage and Speech*, **12**, 156–66.

WRIGHT, P. (1970). The immediate constituents of active and passive sentences. In D'ARCAIS, G. and LEVELT, W. (Eds.), *Advances in Psycholinguistics*. Amsterdam: North Holland Co.

APPENDIX

Experimental sentences: the phrases in italic correspond to the correct answers in the Question conditions

1. The *qualified managing director* was recently sensibly appointed by the expanding successful company.
2. The eminent world-famous physician was normally thoughtfully respected by *the high-ranking distinguished official*.
3. The *exhausted despondent travellers* were frequently cheerfully encouraged by the ambitious intrepid explorer.
4. The abnormal misshapen fingerprint was *skillfully patiently magnified* by the curious forensic scientist.
5. The diligent inquiring policeman was viciously noisily obstructed by the *malicious quarrelsome teenager*.

6. The abortive ill-fated explosion was finally entirely regretted by the *scurrilous marauding terrorists.*
7. The flourishing productive enterprise was *foolishly completely mechanized* by the meddlesome graduate engineer.
8. The delicious thirst-quenching lemonade was ceaselessly volubly demanded by the *excited talkative visitors.*
9. The *rapacious immoral trespasser* was cynically thoroughly delighted by the forbidden underground rendezvous.
10. The beautiful cosmetic assistant was *thankfully speedily detected* by the hurrying observant customer.
11. The *miserable shipwrecked commodore* was cruelly painfully buffeted by the merciless relentless elements.
12. The impressive colourful tradition was *stupidly thoughtlessly abolished* by the unthinking power mad dictator.
13. The abandoned derelict edifice was *brutally wantonly vandalized* by the destructive ignorant delinquents.
14. The befuddled hesitant coroner was constantly properly instructed by the *affable lenient magistrate.*
15. *The majestic important occasion* was willingly cleverly televized by the creative Mexican producer.

Quarterly Journal of Experimental Psychology (1971) 23, 214–224

FUNCTIONAL STRUCTURE IN SENTENCES: A PERFORMANCE ANALYSIS

R. A. KENNEDY AND A. L. WILKES
University of Dundee

Two experiments are reported which attempt to define the groupings of component words within sentences which subjects have committed to memory. The structural groupings are indexed by judgement latencies for pairs of words and these serve as the input matrix for a hierarchical clustering (HC) analysis. It is concluded that when subjects make judgements concerning the forward order of pairs of words, the latencies imply the presence of a hierarchical organization. Although the tree structures obtained do not follow in any detail the surface structures of the sentence types in either experiment, nonetheless when constituent analysis indicates no difference it is accompanied by identical performance structures, and when a surface distinction is called for, an appropriate difference is found in the tree diagrams produced by cluster analysis. Deep structure differences involving the rearrangement of component words are not found in the hierarchical structure subjects imposed. The pausing patterns followed by subjects when reading the sentences are shown to relate to the structural diagrams generated by the HC analysis.

Introduction

A comprehensive structural description of a sentence is ideally one that would specify the relationship of all component words, one with another. The types of relationship would cover such features as "modifier", "subject of", and "object of", indicating in each case the relevant words involved. Some form of transformational grammar may provide this information at the level of competence, and, it is sometimes argued, will also provide a predictive description at a performance level. Such measures, as word "chunking" in learning and recall, encoding units in speech perception, and pause boundaries in reading have been interpreted in this way. The argument has been extended in studies concerned with times to recognize transformations of standard sentences (e.g. Miller, 1962; Miller and McKean, 1964), and experiments on the relative difficulty of processing sentences of varying complexity defined in terms of transformational history (Mehler and Carey, 1967; Rohrman, 1968). Generally, it is thought that some additional parameters will be needed to bridge the areas of competence and performance, but in the main these have been seen as dealing with capacity limitations rather than with the basic descriptive units. In contrast, the experiments reported in the present paper are entirely concerned with an analysis, at a performance level, of the units subjects employ in a recognition task with sentence material.

A retrieval latency task (Kennedy and Wilkes, 1968) has provided some evidence

on the extent to which structural relationships influence performance when subjects are operating on previously memorized sentences. The "latency profiles" which were obtained when subjects responded to a particular word by giving the word immediately following, suggested priority of access to words occupying certain positions (Wilkes and Kennedy, 1969; Kennedy and Wilkes, 1970). Retrieval across major constituent boundaries was retarded, relative to other positions in a sentence. There was also a significant retardation in retrieval latency at other points which had no clear dependence on constituent structure. The location of pauses, indexing the particular grouping pattern subjects adopted when reading sentences, emerged as a good predictor of retarded latencies (Wilkes and Kennedy, 1970*a*, *b*).

In another experimental situation, in which subjects responded to probe words simply indicating if they were present or not in a sentence, the latency profiles could not be systematically related either to syntax or to reading pauses; however, subject and object nouns were associated with rapid latencies (Kennedy and Wilkes, 1969). These techniques of recording response times as subjects process words within a sentence thus appear to define certain of the structural and semantic features relevant to performance in the experimental tasks.

An obvious limitation of the methods discussed so far is their inability to provide measures of the structural features that contribute to retrieval strategy when non-adjacent words are involved. In an attempt to define these features the present experiment made use of a technique which provided an exhaustive sampling of word-pair comparisons within short sentences of varying grammatical structure. The experimental method involved timing subjects when deciding whether visually presented pairs of words taken from a previously learned sentence were in their original order of occurrence; subjects pressed one key to indicate a correct order and a second key to indicate a reversed order. If it is assumed that a subject can make a decision about word order more rapidly for one word pair than for another, the resulting set of differences in levels of relatedness may be used to index structural features in the way that simple response latencies were used previously. Providing the set of relatedness measures shows sufficient internal consistency to be categorized in some way, it is possible to compare such behavioural structural descriptions with alternative, syntactic, descriptions of the same sentences.

When a subject memorizes a sentence the structural and semantic information that can potentially be abstracted is defined by both the deep structure and the surface constituents with their higher order constructions. These inter-word relationships can be seen as leading to each word in the sentence being "cross-indexed" with other words. For example, an adjective may only relate to the noun it qualifies, whereas a subject noun may relate not only to the words within the subject phrase but also to words in phrases later in the sentence. In summary, the general hypothesis under consideration is that judgement latencies to any pair of words, should relate to the degree of common reference within this cross indexing shared by the two words. In terms of the task demanded of the subject in the present situation, contributory components of this response latency are the accessibility of individual words within the sentence—already shown to vary with

both structural and semantic properties—and a possible additional facilitation or retardation relating to the particular properties of the pair.

The word-order decision latencies obtained in this way can not only be used to isolate word groupings relevant to a subject's performance, but also can in principle be analysed for evidence of a hierarchical combination, and for this reason the method of data categorization chosen was that of hierarchical clustering (HC) analysis (Johnson, 1967). A similar procedure has been used by Levelt (1970) in the analysis of "chunking" patterns in the processing of sentences. In Experiment I two contrasting sentence types of the form *He is happy to perform* and *He is heavy to lift* were compared in terms of the hierarchical clusters they elicited. These two sentence types differ in deep structure while possessing identical surface structure (Mehler and Carey, 1967). In Experiment II sentence types of the form *Many sounds are echoing voices* and *Many farmers are making hay* were contrasted. These sentence types differ in both deep and surface structure (Rohrman, 1968).

Experiment I

Method

Materials

Two sets of five sentences based on those of Mehler and Carey (1967) were used.

Type I	He is happy to perform
	They are reluctant to come
	She is hesitant to talk
	Many are eager to buy
	It was ready to fall
*Type II**	He is heavy to lift
	They are troublesome to employ
	She is dangerous to excite
	Many were easy to please
	It was awkward to move

Procedure

Two groups of ten subjects were assigned the five sentences of a given type. Within each group a subject was shown a particular sentence typed on a card and asked to read it; the card was then concealed and he was asked to recall the sentence. This procedure was continued until a criterion of three consecutive correct recalls had been reached. The subject was then called upon to respond to a random series of all 20 pairs of words from the sentence by pressing a key with the right hand to indicate that the words were in their correct order, or another with the left hand to indicate they were reversed. The word pairs were presented visually for 2 sec and the interval between pairs was 10 sec. The remaining four sentences were treated in the same way and sentence order within each group of subjects randomized.

Response latencies to the presented word pairs were automatically timed in msec using a Solartron data logging system. Testing was carried out in an acoustic booth remote from the recording equipment, and an experimental session lasted about 30 min.

* It has since been pointed out that the sentence *he is heavy to lift*, unlike the other type II sentences, has no paraphrase of the form *to* VN *is heavy*. Also, *it was awkward to move* is a type II sentence in only one of its two readings.

Results and Discussion

The HC analysis consists of the following steps. Considering an $n \times n$ matrix of distance measures, D: (1) the smallest is located and taken as the basis of the first cluster. If the measure is $D(i, j)$ the cluster involves i and j and is given a "value" equal to $D(i, j)$. (2) If rows i and j are not adjacent then the matrix is adjusted by reordering the sequences of rows and columns i to n, to bring rows i and j together. (3) Entering the matrix along rows i and j, the pairs of measures (i and j, 1) to (i and j, n) are examined in turn, replacing each pair by a single value equal either to the larger of the two (diameter method) or the smaller (connectedness method). (4) The procedure is then repeated for columns i and j. (5) This process produces a new matrix $(n\text{-}1) \times (n\text{-}1)$ the diagonal of which is set at zero. (6) A further selection of the smallest distance measure is then made and this constitutes the basis of the next cluster into which a previous cluster may or not be incorporated. The total sequence is then repeated, reducing the dimensionality of the matrix by one at each iteration until all values fall into a single cluster.

Data analyses were based entirely on the word pairs presented in their correct order since these pairs provided the most direct test of the hypotheses under investigation. For both sentence types all subjects reached criterion on the first three learning trials and error rates in the word-order task did not differ reliably, the overall mean being 8·8%.

The input matrices for both sentence types are shown in Table I in which the cell entries are means for correct forward response latencies averaged over ten subjects and five sentences. Following the assumptions concerning response latency and relatedness measures outlined earlier an HC analysis was conducted and the tree structures are shown in Figure 1. Mean response latency for sentences of Type I was 1156 msec and for sentences of Type II was 1061 msec, this difference was not significant ($F = 1{\cdot}04$, $df = 1, 18$).

In the sentence *he is happy to perform* the word *he* functions as the subject of *perform*, whereas in the sentence *he is heavy to lift*, *he* is the direct object of *lift*. This distinction goes unmarked in the constituent description of the surface structure of the sentences but is drawn in their related deep structures (Mehler and Carey, 1967).

TABLE I

Input matrix of latencies derived from forward word-order comparisons in Type I and Type II sentences (Experiment I)

	Type I						Type II				
	1	2	3	4	5		1	2	3	4	5
1		929	1064	1152	1095	1		859	967	1082	1030
2			1138	1314	1449	2			975	1240	1291
3				1160	1297	3				1025	1138
4					1010	4					1001
5						5					

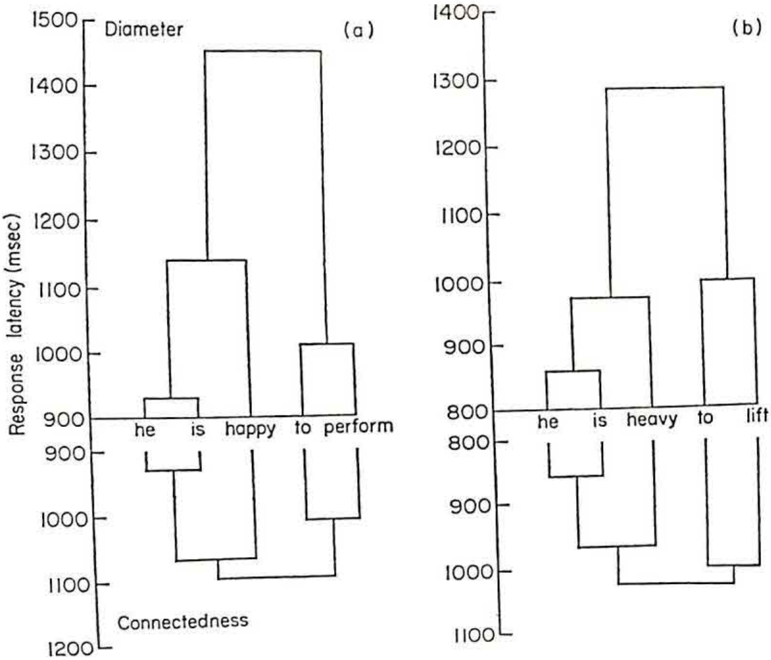

FIGURE 1. Hierarchical cluster analysis of Type I (a) and Type II (b) sentences (Experiment I) derived from forward word-order comparison latencies. Diameter and connectedness methods of analysis are shown separately.

The surface structure for both sentence types is given by (*He* (*is* (*happy* (*to* (*perform*))))) and (*He* (*is* (*heavy* (*to* (*lift*))))), Katz and Postal (1964, p. 37). It has been pointed out, however, that *to be* has the principal function of marking tense, mood and aspect in surface structure and need not be a constituent of deep structure. "Grammars of English which introduce *be* into the deep structure of sentences will necessarily contain many (additional) transformational rules for the deletion of this element in a variety of positions in surface structure" (Lyons, 1968, p.323). The major deep structure distinctions between sentences of Types I and II are unaffected by this but it does introduce ambiguities into the constituent analysis of surface structure. An alternative surface structure description that would allow for the copulative function of *be* is ((NP)-(C)-(NP)), i.e. a three-way partitioning. Since HC analysis operates on binary partitioning the data present no direct information on this issue. The cluster hierarchies derived from both methods of analysis do not differ from one sentence type to the other and hence do not mark their deep structure distinctions. The performance structures do however consistently differ from the results of a constituent analysis that includes *be* with the predicate. The principal departure relates to the three-word groupings ((*he* (*is*)) (*heavy*)) and ((*he* (*is*)) (*happy*)) This form of organization is not a necessary consequence of the experimental procedure since, in unpublished work in our laboratory using unstructured lists of five words the cluster 1–5 is commonly found, together with a reordering of items. It follows, therefore, that subjects in the present experiment, while operating on both sentence types using a common

structuring principle, did not exlusively use the structure suggested by a binary form of constituent analysis.

The most reliable indication that a data matrix can be validly described by a hierarchical structure is that the two methods of analysis (diameter and connectedness) should yield identical or near identical results. Reference to Figure 1 shows that this condition is met. This argument stems from the assumption that when merging two rows or columns the decision to replace each pair of cell entries by either the larger or smaller value should make no material difference to the ordinal relationship between the new cell entry and all other cells in the matrix.

In conclusion, it appears there are two ways of describing the sentences used in Experiment I. The first is in terms of the various syntactic relationships as summarized in deep and surface descriptions, and the second is in terms of the word groupings underlying task performance. The relation of these two descriptive procedures is discussed in more detail below but there is obviously no evidence that the difference in deep structures of the two sentence types used led to a reordering of words or groups of words to meet more directly the logical relationships implied.

Experiment II

Since the surface structures of the two sentence types in Experiment I were identical it could be argued that the derived tree diagrams did not in fact relate to the grammatical relationships present in the test sentences but to some specific principles of organization peculiar to this task and method of analysis. The second experiment set out to examine the effects on performance of a difference in both surface and deep structure in order to provide comparative data when task and mode of analysis remained unchanged but additional grammatical features were introduced.

Method

Materials

Two sets of five sentences based on those of Rohrman (1968) were used.

Type I	Many sounds are echoing voices Most canaries are singing birds These observers are sailing men Some plants are flowering shrubs Those animals are growling lions
Type II	Many farmers are making hay Most people are raising flowers These students are taking notes Some visitors are naming trees Those men are shoeing horses.

Subjects

Twenty subjects were employed, all volunteers, drawn from a Summer School for American students held in the University of Dundee.

Procedure

Each of two groups of ten subjects were assigned the five sentences of a given type, and the design and execution of the experiment followed exactly that described for Experiment I. As with certain sentences in Experiment I an alternative, if unlikely, reading may be assigned

TABLE II

Input matrix of latencies derived from forward word-order comparisons in Type I and Type II sentences (Experiment II)

	Type I						Type II				
	1	2	3	4	5		1	2	3	4	5
1		1224	1280	1368	1479	1		906	1095	1192	1223
2			1335	1452	1502	2			1045	1201	1193
3				1357	1514	3				1006	1280
4					1401	4					1164
5						5					

to the Type I sentences in isolation (reading "voices," "birds," "men," etc., as the direct objects of "echoing," "singing," "sailing," etc). The fact that subjects experienced only one set of sentences, however, made such interpretations highly unlikely.

Results and Discussion

All subjects reached criterion on the first three learning trials. Overall mean error rate, which did not differ significantly for the two sentence types, was 7·8%. The input matrices of correct forward latencies are given in Table II. The cell entries are averages across ten subjects and five sentences. Figure 2 shows the tree structures defined by the HC analysis. The absolute levels of response latency for sentences of Type I was 1392 msec and for sentences of Type II was 1127 msec, this difference was not significant ($F = 3{\cdot}17$, $df = 1, 18$, $P > 0{\cdot}05$).

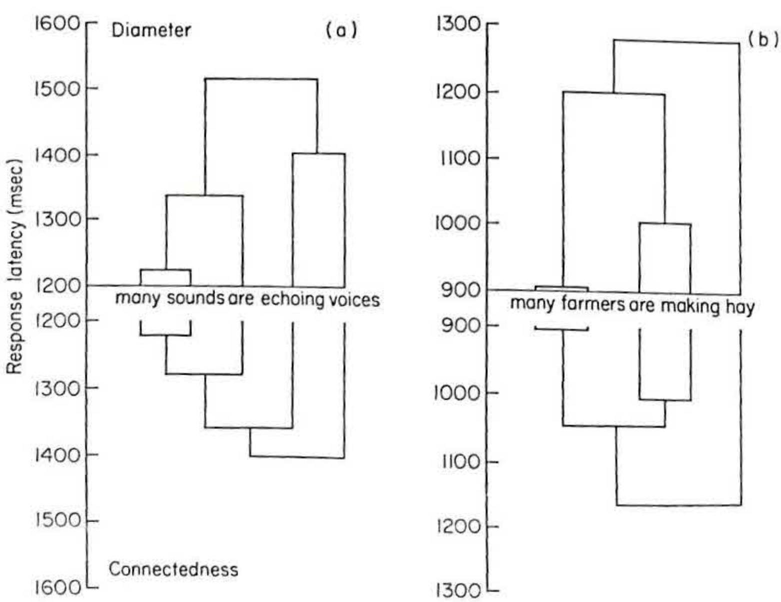

FIGURE 2. Hierarchical cluster analysis of Type I (a) and Type II (b) sentences (Experiment II) derived from forward word order comparison latencies. Diameter and connectedness methods of analysis are shown separately.

The differences in surface structure focus on the additional "depth" of sentences of Type II which has a verbal auxiliary. The surface structures† are respectively, Type I: ((*Many* (*sounds*)) (*are* (*echoing* (*voices*)))); and Type II: ((*Many* (*farmers*)) (*are* (*making*)) (*hay*)). The deep structure relationships in Type II sentences are similar to those in the surface structure. In the case of the Type I examples the deep and surface structure differ profoundly (Rohrman, 1968).

The structures abstracted from the results of subjects' forward comparisons differed for the two sentence types. In respect of the main surface structure difference, namely the presence of an auxiliary verb in Type II, it can be noted the HC analysis clearly illustrates this distinction. The resulting tree diagrams do, however, characterize meaningful word groupings within the sentences without exactly reproducing any expected surface structure. The deviation from surface structure resides in the incorporation of the "verbal unit" into the subject phrase. Clear subject, "verb", and object word groupings can be identified in both sentence types: the major differences arise only when the hierarchical organization is considered.

Once again the deep structure differences between sentence types have not produced any reordering of the component words; however, the structurally more complex type tends to be associated with longer latencies. The diameter and connectedness methods of analysis yield identical solutions for Type II sentences but not for Type I, a finding which casts some doubt on the validity of the hierarchical structures obtained in this particular case. This point is dealt with later in the discussion.

In summary, the two experiments reported allow the following conclusions: (1) Using short sentences, when subjects make judgements concerning the forward order of pairs of words, the latencies imply an organization derived only in part from grammatical features. Although the tree structures obtained do not follow in any detail the surface structures of the sentence types in either experiment, nonetheless when constituent analysis indicates no difference it is accompanied by identical performance structures, and when a surface distinction is called for an appropriate difference is found in the tree diagrams produced by cluster analysis. (2) Deep structure differences involving the rearrangement of component words are not found in the hierarchical structure subjects imposed. (3) In its present state HC analysis is restricted by a sensitivity to slight variations in the input matrix, and a lack of known sampling distributions for clusters. The technique does, nevertheless, provide a unique means of extracting relationships from latency data based on multiple word-order decisions. The ultimate validation of these relationships must lie in the predictive generality of the structures they define.

Reading pauses have been used in earlier studies (Wilkes and Kennedy, 1970*a*) to provide an independent index of the word groupings subjects employed in a probe retrieval task. For this reason in an attempt to secure some validation of the structuring shown by HC analysis, an examination was made of the pausing patterns used by subjects when reading the test sentences. The method of pause measurement involved writing the square wave output of a voice key on to a pen

† But see earlier discussion of the role of the copula.

recorder using subjects' readings and recalls of the sentences as input. Tapes were played at half speed using a pen recorder paper speed of 10 mm/sec. Inter- and intra-word breaks were distinguished by visual inspection. The mean pause duration between pairs of words (readings and recalls averaged) are shown in Table III.

The first and most direct comparison that can be made between the judgements of word order and the probe retrieval judgements called for in earlier experiments resides in the latencies for adjacent word pairs. In probe retrieval a retarded latency to respond with word $n + 1$ in a sentence, given word n as probe, was best predicted by the presence of a lengthened pause between these positions when the sentence was read, (Wilkes and Kennedy, 1970*a*). While pause-duration and

TABLE III

Mean pause duration (msec) for the sentences of Experiments I and II

		Words			
		1–2	2–3	3–4	4–5
Expt. I	Type I	29	66	152	105
	Type II	18	79	140	60
Expt. II	Type I	120	142	114	133
	Type II	127	88	53	142

latency did not correlate perfectly it was concluded that the conjunction of lengthened pause and retarded latency arose because both were dependent on some underlying structure the subject imposed on the sentence. It can be predicted, therefore, that in the present experiments the structural divisions subjects employed in learning the sentences will be marked in a similar manner, and that, considering only adjacent word pairs, where these straddle a pause-defined boundary, the latency for judging them to be in their correct order will also be retarded. Reference to Table III (Experiment I) indicates that it is the word pairs occupying positions 3 and 4 in both sentence types that meet this criterion. Cross reference to Table I provides evidence that of the four possible adjacent pair comparisons it is this one which is processed slowest in both conditions. In Experiment II the relationship does not hold for Type I sentences, but does hold for Type II. Thus the latencies in probe retrieval and in word-order comparisons of adjacent words both show an association with pausing strategy; that is, each relates to an independent measure of how subjects appear to structure the sentence. Unlike adjacent probe latencies the word-order decision latencies can be analysed as a total set, providing structural tree diagrams for all of the sentence types. It remains therefore to look for correspondences between word groupings defined by HC analysis and those independently defined by pauses.

In Experiment I the pause records show a clear break between the third and fourth words of both sentence types. This finding is highly consistent, appearing in eight of the ten sentences (in the two exceptions the second longest pause was

found in this location). If this, the longest pause, is taken as an index of a subject's decision as to the primary division in the sentences, it can be seen this is also the location of what might be called the "immediate constituent break" in the tree diagrams of Figure 1.

In Experiment II for Type II sentences the longest pause again marks the primary division in the cluster diagrams. The pausing pattern is consistent, holding for four of the five sentences. For Type I sentences no clear pause distribution can be identified: there is wide inter-sentence variability, and predictably it is this sentence type which yields a hierarchical structure of doubtful validity.

In conclusion, the results presented accord with earlier findings that subjects organize sentences for retrieval and recognition tasks using groupings which are difficult to relate unambiguously to syntax. It has been pointed out elsewhere that the syntactic analysis of the sentence types used is by no means straightforward. If a tripartitioning of surface structure is adopted then, by definition, the present methods of analysis could not directly reproduce this. The general tendency to isolate the copula with an eventual assimilation into the subject phrase could be given some linguistic justification since the properties of the subject phrase serve to determine the form that *be* takes when inserted transformationally. Thus, it might be argued that deep structure features have been reflected in the structural diagrams obtained. Such an argument, to be consistent, must neglect the fact that the auxiliary and main verb in Experiment II Type II sentences are treated in a similar manner.

The particular groupings of words shown by HC analysis highlight a difficulty in psycholinguistic research which stems from the fact that the theory from which predictions of surface structure may be made is a theory of linguistic competence. In many cases it has been possible to make use of such a theory as a performance model, but the present results at least suggest that a complete account in psychological terms will need to draw more directly on evidence relating to the units of organization which appear relevant to a subject for any particular task. Hopefully, given a reasonable variety of experimental situations, it may eventually be possible to examine the precise interrelatedness of performance data and theories of linguistic competence.

This work is supported by a grant from the Social Science Research Council. We gratefully acknowledge the assistance of Mrs R. Ellen in the collection and analysis of the data.

References

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, **32**, 241–54.

Katz, J. and Postal, P. (1964). *An Integrated Theory of Linguistic Descriptions.* Boston, Mass.: M.I.T. Press.

Kennedy, R. A. and Wilkes, A. L. (1968). Response-times at different positions within a sentence. *Quarterly Journal of Experimental Psychology*, **20**, 390–4.

Kennedy, R. A. and Wilkes, A. L. (1969). Analysis of storage and retrieval processes in memorizing simple sentences. *Journal of Experimental Psychology*, **80**, 396–8.

Kennedy, R. A. and Wilkes, A. L. A comparison of retrieval and location operations in active and passive sentences. *British Journal of Psychology*. In Press.

Levelt, W. J. M. (1970). Hierarchical chunking in sentence processing, *Perception and Psychophysics*, **8**, 99–102.

Lyons, J. (1968). *Introduction to Theoretical Linguistics*. London: Cambridge University Press.

Mehler, J. and Carey, P. (1967). Role of surface and base structure in the perception of sentences. *Journal of Verbal Learning and Verbal Behaviour*, **6**, 335–8.

Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, **17**, 748–62.

Miller, G. A. and McKean, K. (1964). A chronometric study of some relations between sentences. *Quarterly Journal of Experimental Psychology*, **16**, 297–308.

Rohrman, N. L. (1968). The role of syntactic structure in the recall of English nominalizations. *Journal of Verbal Learning and Verbal Behaviour*, **7**, 904–12.

Wilkes, A. L. and Kennedy, R. A. (1969). Relationship between pausing and retrieval latency in sentences of varying grammatical form. *Journal of Experimental Psychology*, **79**, 241–45.

Wilkes, A. L. and Kennedy, R. A. (1970*a*). Response retrieval in active and passive sentences. *Quarterly Journal of Experimental Psychology*, **22**, 1–8.

Wilkes, A. L. and Kennedy, R. A. (1970*b*). The relative accessibility of list items within different pause defined groups. *Journal of Verbal Learning and Verbal Behaviour*, **9**, 197–201.

Received 3 June 1970

Quarterly Journal of Experimental Psychology (1971) 23, 225–233

THE INFLUENCE OF VISUAL TEXTURE DENSITY GRADIENTS ON RELATIVE DISTANCE JUDGEMENTS

COLIN V. NEWMAN
University of Birmingham

A hypothesis derived from J. J. Gibson's psychophysical theory of space perception was tested. Subjects made monocular relative distance judgements by moving a marker to the apparent physical mid-point between two other fixed markers which were placed on a surface along the subjects' line of sight. Judgements were significantly influenced by the texture density gradients of stimulation derived from the surface over which they were made.

Introduction

In his theory of visual space perception, Gibson (1950, 1959, 1961) emphasizes the role played by higher-order variables of stimulation present in the retinal array. In particular, texture density gradients are suggested as important sources of information about depth. Though a number of experimenters have investigated certain aspects of Gibson's psychophysical hypothesis (Epstein and Park, 1964), few studies have considered the influence of texture variables on judgements of *distance*. In a number of studies, by Gibson (1947), Weinstein (1950, 1957), Smith (1958), Sonoda (1961), Wohlwill (1962, 1965) and Newman (1969), the influence of texture and perspective gradients on distance judgements within a two-dimensional, pictorial scene has been considered. From these experiments it is difficult to draw conclusions relating to the wider issue of space perception in the three-dimensional world of objects and surfaces since, as both Gibson (1954) and Hochberg (1962) indicate, it cannot be assumed that the information about depth provided by a picture is exactly equivalent to that provided in the real world even when the observer is motionless and monocular.

When three-dimensional materials have been used by experimenters in the past, they have not tested satisfactorily the influence of surface texture on distance judgements and have reached unjustified conclusions. Whitehouse and Gruber (1957) studied bisection judgements of the distance between two markers but limited their study to a gross comparison of a texture and a textureless condition. In a number of similar experiments, Wolhwill (1963*a*, *b*, 1964) found that surface texture had little consistent influence on bisection judgements and therefore he concluded that Gibson may have placed undue stress on the importance of the texture variable. It can be questioned whether Wohlwill's experiments provide an adequate test of Gibson's view. All the surfaces over which bisection judgements were made provided subjects with identical *gradients* of texture density, even though they varied in the amount of information defining the gradients. The number of texture

elements per unit area of surface was constant over each individual surface. Surfaces varied only with regard to the number and regularity of the distribution of the texture elements. Gibson (1959) argues that an observer's judgement of the length of one stretch of ground relative to another depends on the number of transitions or texture units in the stretch, relative to the number in the other stretch. Therefore, when a subject is asked to indicate the mid-point of a surface, he will make this particular type of relative distance judgement by attempting to equate the number of texture units beyond the mid-point with the number of similar units in front of it. If this strategy of judgement were followed on Wohlwill's surfaces, an objectively correct judgement of the physical mid-point would be achieved on every surface, since there was always an equal number of texture units in opposive halves of the same surface.

In order satisfactorily to investigate the influence of texture density gradients on bisection judgements, the packing density of elements comprising the texture must be progressively varied along the surface. The apparent mid-point specified in the texture density gradient will then be different from the objective mid-point determined by the physical characteristics of the apparatus. Artificially deformed texture density gradients have not been used in any previous experiments in which subjects have made relative distance judgements. The present experiment employed test surfaces of this design, in order to discover whether relative distance judgements are influenced by texture density gradients of stimulation.

Experiment I

Method

Apparatus

Figure 1 shows a schematic picture of the apparatus.

Dexion girders were bolted together to form a rectangular frame ("G" in Figure 1) mounted horizontally, 2·5 ft above the floor. A large viewing box (B), made of unpainted hardboard, was secured to one end of this frame. The subject could look into this box through a monocular viewing aperture (A), 0·75 in. in diameter, located at the centre of the front side of the viewing box. Adjustable buffers (C) were provided to support the subject's head at the temples and thus prevent any lateral movement of the eye. The side of the viewing box opposite to that carrying the viewing aperture consisted of a reduction screen (D) with a single aperture, slightly to the right of the subject's direct line of sight. A sliding door (E) was available to cover this aperture. With the door open, the subject could view the centre portion and ends, but not the sides, of a brightly-lit test surface (K) placed on the Dexion framework of the apparatus with one end flush with the floor of the viewing box. Vision beyond the end of the test surface was restricted by a white screen (I), mounted vertically, 1 ft beyond its end. The test surface (K) could easily be removed from the apparatus. Several interchangeable test surfaces were prepared with different textures, as described later. All were cut to the same dimensions from sheets of hardboard. Along the centre of the subject's line of sight were placed two markers (F and J), one near to and the other further from the subject's viewing position. These markers defined the distance to be bisected. The far marker (J), placed almost at the horizon end of the test surface, was a small, blue, vertically mounted, cardboard rectangle, 0·87 in. wide by 1·22 in. tall. The near marker (F), a small wooden cylinder 1 in. tall and 0·50 in. in diameter, was placed towards the near end of the test surface 1 ft from its junction with the floor of the viewing box. During the experiment, the distance between the two end markers could be bisected by adjusting the position of a bisection marker (H). This consisted of a small wooden cylinder, 0·38 in. tall by 0·50 in. in diameter, mounted on a 1·10 in. diameter magnetic

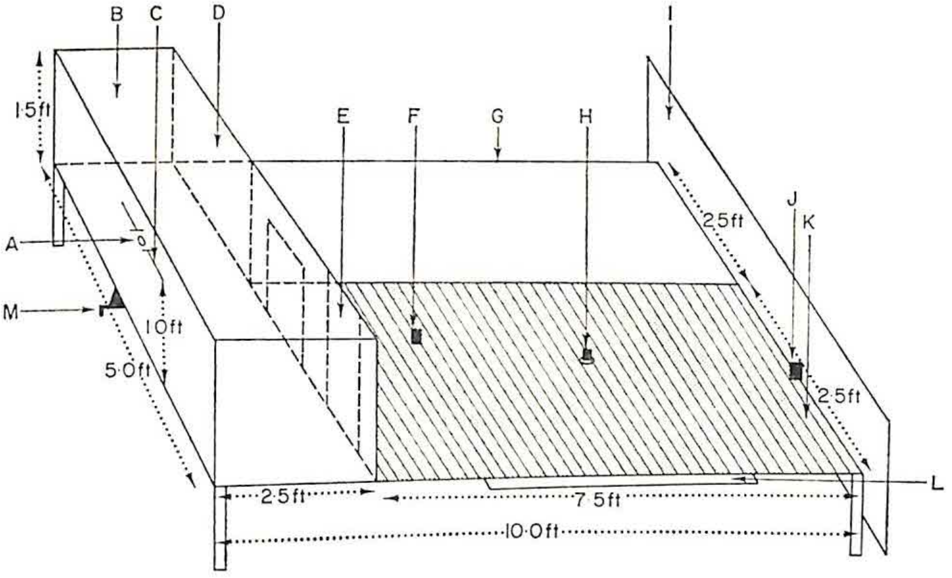

FIGURE 1. Schematic picture of the apparatus. A—Monocular viewing aperture; B—viewing box; C—adjustable buffers; D—reduction screen with aperture; E—sliding door; F—near marker; G—rectangular frame of apparatus; H—distance bisection marker; I—white screen; J—far marker; K—test surface; L—measuring scale; M—winding handle.

steel disc. This bisection marker was placed on the test surface above a powerful button magnet mounted on tracks and held immediately under the test surface. By winding a handle (M), mounted to the apparatus below the viewing box, the subject could vary the position of the button magnet and hence move the bisection marker along the length of the surface between the end markers. The subject's judgements of the mid-point between the two end markers could be read from a scale (L), mounted by the side of the test surface and visible only to the experimenter.

The test surfaces

Six interchangeable test surfaces painted with a matt finish were used. One was white and the remaining five all had alternate latitudinal black and white stripes painted across them. On two surfaces all the stripes were regular in width. The other three surfaces bore stripes that became progressively wider from one end of the surface to the other. The six surfaces were as follows.

Plain white (P.W.). This was painted white all over, and had no visible microstructure.

Large regular textured (L.R.T.). Covered with 45 stripes all 2·0 in. wide.

Small regular textured (S.R.T.). Covered with 63 stripes all 1·43 in. wide.

Fast decreasing textured (F.D.T.). Covered with 75 stripes. The stripe nearest to the subject's viewing position was 2·0 in wide, each successive stripe being 1/70th narrower than the previous one, making the final horizon stripe 0·67 in. wide.

Slowly decreasing textured (S.D.T.). Covered with 63 stripes. The stripe nearest to the subject's viewing position was 2·0 in. wide, each successive stripe being 1/90th narrower than the previous one, making the final horizon stripe 0·98 in. wide.

Slowly increasing textured (S.I.T.). Covered with 63 stripes. The stripe nearest to the subject's viewing position was 0·98 in. wide, each successive stripe being 1/90th wider than the previous one, making the final horizon stripe 2·0 in. wide.

On all surfaces an odd number of stripes was used, to allow both the nearest stripe to the subject and the furthest horizon stripe to be painted black.

Procedure

Ninety-six subjects, male and female, undergraduates and postgraduates, took part in this experiment. Their ages ranged from 17 to 37 years. All the subjects had normal or corrected normal vision. They were divided randomly into 6 experimental groups with 16 subjects in each, but they were tested individually, making distance judgements on first one, then another of the test surfaces selected to meet the criteria given below. Subjects in each experimental group made judgements on the following surfaces.

Group 1: F.D.T. first, then L.R.T. (gradients differ, but initial stripe equal in width on both surfaces).

Group 2: S.D.T. first, then S.R.T. (gradients differ, but equal number of stripes on both surfaces).

Group 3: L.R.T. first, then P.W. (control comparison of regular texture and no visible texture).

Group 4: S.R.T. first, then L.R.T. (control comparison of surfaces with small or large stripes but constant gradient).

Group 5: P.W. first, then L.R.T. (control to investigate any order of presentation effects by comparison with Group 3).

Group 6: S.I.T. first, then S.R.T. (gradients differ, cf. Group 2, but equal number of stripes on both surfaces).

Each subject was asked to look through the monocular viewing aperture at the first surface on which he was to make his distance judgements. The near and far markers were pointed out to him and the manipulation of the distance bisection marker practised. The experimenter postioned the bisection marker close to the near marker, and instructed the subject to adjust the position of the bisection marker until it was equidistant between the near and far markers. The stripes on the test surfaces were not mentioned. If the subject asked if he could deliberately count the stripes, the experimenter indicated that this was not permitted. In practice, it was impossible to obtain an accurate count of the stripes, when viewed from the subject's position. If any subject had attempted a deliberate stripe-counting strategy when making his bisection judgements, this would have become apparent to the experimenter both by the time the judgement would have taken and by the way the bisection marker would have had to be moved along the surface. After the subject had made his first judgement, he was questioned to ensure that the instructions had been fully understood and obeyed. The experimenter next closed the reduction screen door and recorded the subject's judgement. If the bisection marker were adjusted beyond the objective mid-point, this was scored as a positive deviation from the mid-point. A negative deviation was recorded when the bisection marker was placed nearer than the mid-point. The subject repeated his bisection judgement three more times, the starting position of the marker being so arranged that the subject's judgements were made in the conventional ascending, descending, descending, ascending sequence. The subject's point of subjective equality was calculated as the mean of all four judgements. The whole procedure was repeated with the second test surface used for each subject.

At the end of the experiment, the subject was questioned about the nature of the surfaces he had seen, in order to discover if he had been able to perceive correctly the physical nature of the stripes.

Results

On the hypothesis that texture influences relative distance judgements, the mean distance bisection scores on the six test surfaces will fall in the sequence indicated below. Furthermore, it is possible to make quantitative theoretical predictions of the subjective mid-points of the variously textured surfaces. The following mean deviation scores from the physical mid-point of each surface can be predicted if

subjects adjust the bisection marker so as to equate the number of texture units beyond the mid-point, with the number of similar units in front of it.

$$\begin{array}{cccccc} \text{F.D.T.} > & \text{S.D.T.} > & \text{L.R.T.} = & \text{S.R.T.} = & \text{P.W.} > & \text{S.I.T.} \\ +9{\cdot}29 \text{ in.} & +6{\cdot}18 \text{ in.} & 0{\cdot}0 \text{ in.} & 0{\cdot}0 \text{ in.} & 0{\cdot}0 \text{ in.} & -5{\cdot}43 \text{ in.} \end{array}$$

Table I shows the mean distance bisection scores for all subjects. Also given are the results of related sample t tests comparing the subjects' bisection scores on the first surface with their bisection scores on the second surface. Intersubject product-moment correlation coefficients for comparisons of these same bisection scores are given in the final column of Table I.

TABLE I

Subjects' mean distance bisection scores in inches

	First surface viewed	Mean and S.D. (in.)	Second surface viewed	Mean and S.D. (in.)	Comparison of means by t tests ($df = 30$)	Correlation r ($df = 15$)
Group 1	F.D.T.	11·38 ±5·37	L.R.T.	4·53 ±6·37	4·81 $P < 0{\cdot}001$	0·420 N.S.
Group 2	S.D.T.	9·49 ±4·38	S.R.T.	6·97 ±3·26	3·28 $P < 0{\cdot}01$	0·701 $P < 0{\cdot}01$
Group 3	L.R.T.	3·54 ±3·78	P.W.	4·33 ±4·00	0·86 N.S.	0·564 $P < 0{\cdot}05$
Group 4	S.R.T.	2·52 ±6·02	L.R.T.	3·27 ±6·21	0·79 N.S.	0·830 $P < 0{\cdot}01$
Group 5	P.W.	1·30 ±4·85	L.R.T.	2·20 ±3·09	0·82 N.S.	0·654 $P < 0{\cdot}01$
Group 6	S.I.T.	−0·83 ±5·28	S.R.T.	2·64 ±6·23	6·20 $P < 0{\cdot}001$	0·940 $P < 0{\cdot}01$

The scores for the six, independent groups of subjects on the surfaces initially viewed were compared by a one-way analysis of variance; the main effect of surface used was significant ($F = 36{\cdot}1$, $df = 5, 90$, $P < 0{\cdot}001$). Theoretical predictions of the mean, subjective mid-points of the six test surfaces were advanced earlier. A trend analysis of the regression of the obtained distance bisection scores on these theoretical predictions showed a significant linear relation ($F = 171{\cdot}8$, $df = 1, 90$, $P = < 0{\cdot}001$) and a non-significant departure from linearity ($F = 2{\cdot}1$, $df = 4, 90$, $P > 0{\cdot}05$). This means that the differences between the obtained bisection scores corresponded with the differences betwen the theoretically predicted bisection scores; however, confirmation of this trend does not imply an exact numerical correspondence between the obtained and predicted scores on each individual surface. As seen from Table I, all obtained means are close to their theoretically predicted values, but all are numerically larger. The probable reason for this is discussed later.

Questioning revealed that no subject was able to report correctly the physical difference between the stripes on the two surfaces viewed. Seven of the 96 subjects reported that the test surface stripes were in reality progressively varying in

width along the surfaces. Four of these subjects had, in fact, viewed only regularly stripped surfaces. Forty-three subjects were completely confident that the stripes on both surfaces viewed were all physically equal in width. Twenty-three of these subjects had, in fact, viewed an irregular striped surface. The remaining 46 subjects reported considering the possibility that the surface stripes might have been irregular in width, but they had no way of establishing this. Twenty-five of these subjects had actually viewed an irregular striped surface and 21 had viewed only regular striped surfaces. χ^2 tests showed there was no significant relationship between a subject's expressed opinions about the surface and his distance bisection judgements. No subject reported attempting to allow for uneven stripe width when making his judgements. The results will be further discussed after those of a second experiment have been reported.

Experiment II

With the clearly defined textures used in Experiment I, it is possible that subjects' judgements may have been influenced in an atypical way. Surface textures may have little significant influence on relative distance judgements if the texture density gradients are less clearly defined. Experiment II was carried out as a control experiment to discover whether texture density gradients influence relative distance judgements on a multi-unit, textured surface. A new textured surface made from several thousand independent units was prepared. This surface, the spotted field (S.F.) surface, had a gradient of texture density equal to that on the S.D.T. surface.

Method

Construction of the spotted field (S.F.) surface

This surface was made in such a way that it was impossible to count the texture units from any viewing position. It was constructed from hardboard painted entirely with matt-finish white paint. Faint lines were then drawn latitudinally across the surface to give adjacent stripes, each 1/90th less wide than the preceding stripe. With the surface placed on the apparatus, the nearest stripe to the subject's viewing position was 2·0 in. wide; 63 stripes in all were drawn, giving a final horizon stripe 0·98 in. wide. One hundred and twenty-five black dots, each 0·20 in. in diameter, were placed at random within the lines defining each stripe, giving an approximately even scatter of spots within each stripe. The faint lines were then obliterated, leaving a textured surface of several thousand spots on which there was a progressive increase in the packing density of the spots from the near end of the surface to the horizon end. Ideally, the actual size of each spot should also have been progressively reduced along the surface, but with the drawing instruments available this nicety could not be achieved.

Procedure

Sixteen naïve subjects, in all respects comparable to those tested in the earlier experiment, made distance bisection judgements, exactly as in Experiment I, first on the S.F. surface and then on the P.W. surface.

Results

The mean distance bisection scores were: S.F. 6·54 in. (S.E. $\pm$ 7·09 in.) and P.W. 4·72 in. (S.E. $\pm$ 6·31 in.). The means were not significantly different. The product-moment correlation coefficient for subjects' judgements on the first and second surfaces was $r = 0{\cdot}666$ ($P < 0{\cdot}01$). Using a series of t tests for independent

samples, the mean distance bisection score on the S.F. surface was compared, in turn, with the mean distance bisection scores on each of the surfaces used in Experiment I. Winer (1962) recommends the use of Dunnett's *t* statistic when such comparisons between several treatments and a further single treatment are made. Significance levels of *t* were obtained from the Dunnett tables, but standard errors were derived separately for each pair of means. Using this conservative test of significance, it was found that the S.F. mean score differed significantly from the P.W. score ($t = 2{\cdot}44$, $df = 30$, $P < 0{\cdot}05$) and from the S.I.T. score ($t = 3{\cdot}32$, $df = 30$, $P < 0{\cdot}01$). No other comparison was significant.

In order to facilitate comparison of the obtained mean distance bisection scores in Experiments I and II with the theoretical predictions of these means, the results are shown graphically in Figure 2. The solid line on this graph indicates the quantitative predictions.

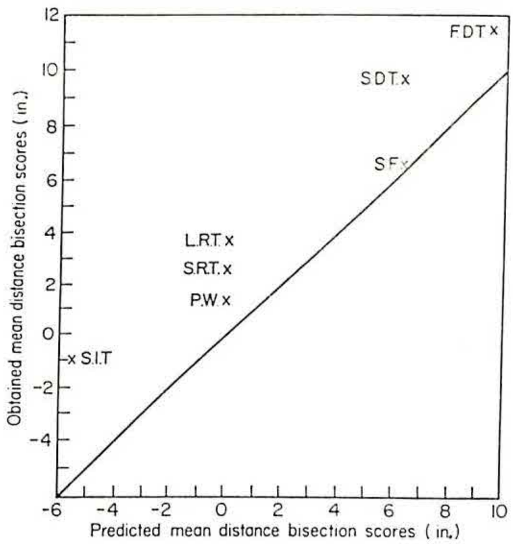

FIGURE 2. Comparison of the obtained mean distance bisection scores of Experiment I and II with quantitative predictions of the means.

Post-experimental questioning revealed that 10 of the 16 subjects agreed that there were more spots per square inch of surface in the distance than near at hand. Mann-Whitney U-tests showed there was no significant difference between the distance bisection scores of these 10 subjects and those of the remaining 6 subjects who held that the dots were evenly distributed along the whole surface. It is probable that no subject appreciated the deformed texture gradient effect as none could understand how they could attempt to allow for the influence of uneven dot distribution when making bisection judgements.

Discussion

The results of Experiment I strongly confirm the experimental predictions, and show that the position of the phenomenal mid-point of a surface was significantly influenced by the gradients of texture density. The mean distance bisection scores

fell in the predicted order of magnitude. The obtained scores on all six surfaces were, however, somewhat higher (more positive in sign) than predicted by objective criteria (see Fig. 2). This shows that on all surfaces, irrespective of their textures, subjects tend to place the bisection marker beyond the mid-point specified by the textural patterns. The same phenomenon has been consistently found in similar distance bisection experiments (Gruber, 1954; Gibson, Bergman and Purdy, 1955; Purdy and Gibson, 1955; Denis-Prinzhorn, 1960; Wohlwill, 1963*a*, *b*, 1964). Wohlwill (1963*a*) terms the effect "over-constancy" and argues that it is judgemental in origin.

The overconstancy of distance judgements is rather less on the S.F. surface of Experiment II than on the surfaces of Experiment I (see Fig. 2). On the S.F., surface, the obtained score (6·54 in.) is only slightly more than the predicted, mean, subjective mid-point (6·18 in.), and is less (though not significantly) than the score obtained on the matching S.D.T. surface. This suggests that the poorly defined S.F. surface had less influence on distance judgements than the better defined striped surfaces. The results have important implications for Gibson's theory, in that they suggest that certain surfaces may provide more "adequate" gradients of texture density than others. Veridical perception may be possible only when certain well-defined textured surfaces are present. Perception may be "directly given" in the way Gibson suggests, only under certain specified conditions. Gibson seems to assume that all surfaces in the everyday environment will provide subjects with sufficiently well-defined texture density gradients to permit veridical judgements of distance. This criticism of Gibson should not detract from the main finding that the higher order stimulus variable, the texture density gradient, has been found to exert a significant influence on relative distance judgements.

The experiments reported in this paper were carried out at the University of Leicester as part fulfilment of the requirements for a Ph.D. degree, supervised by Professor S.G.M. Lee. I held a Social Science Research Council Studentship at the time. The help is acknowledged of Dr Philip M. Levy for his advice on the statistical analysis of the results.

References

Denis-Prinzhorn, M. (1960). Perception des distance et constance des grandeurs (étude génétique). *Archives de Psychologie*, **37**, 181–309.

Epstein, W. and Park, J. (1964). Examination of Gibson's Psychophysical Hypothesis. *Psychological Bulletin*, **62**, 180–96.

Gibson, E. J., Bergman, R. and Purdy, J. (1955). The effect of prior training with a scale of distance on absolute and relative judgements of distance over ground. *Journal of Experimental Psychology*, **50**, 97–105.

Gibson, J. J. (1947). Motion picture testing and research. *A.A.F. Aviation Psychology Research Report* No. 7.

Gibson, J. J. (1950). *The Perception of the Visual World*. Boston: Houghton Mifflin.

Gibson, J. J. (1954). A theory of pictorial perception. *Audio Visual Communications Review*, **2**, 3–23.

Gibson, J. J. (1959). Perception as a function of stimulation. In Koch, S. (Ed.), *Psychology, a Study of a Science*, Vol. 1. New York: McGraw-Hill.

Gibson, J. J. (1961). Ecological optics. *Vision Research*, **1**, 253–62.

Gruber, H. E. (1954). The relation of perceived size to perceived distance. *American Journal of Psychology*, **67**, 411–26.

HOCHBERG, J. E. (1962). The psychophysics of pictorial perception. *Audio Visual Communications Review*, **10**, 22–54.

NEWMAN, C. V. (1969). Children's size judgments in a picture with suggested depth. *Nature, London*, **223**, 418–20.

PURDY, J. and GIBSON, E. J. (1955). Distance judgments by the method of fractionation. *Journal of Experimental Psychology*, **50**, 374–80.

SMITH, O. W. (1958). Judgments of size and distance in photographs. *American Journal of Psychology*, **71**, 529–38.

SONODA, G. (1961). Perceptual constancies observed in plane pictures. In, AKISHIGE, Y. (Ed.), *Experimental researches on the Structure of the Perceptual Space*. Fukuoka, Japan: The Faculty of Literature of Kyushu University, Bulletin No. 7, 199–228.

WEINSTEIN, S. (1950). An analysis of the effects of texture-gradient, position in the field and apparent size, on the perception of depth. Unpublished M.A. thesis, University of New York.

WEINSTEIN, S. (1957). The perception of depth in the absence of texture-gradient. *American Journal of Psychology*, **70**, 611–5.

WHITEHOUSE, J. M. and GRUBER, H. E. (1957). The effects of surface texture and binocular disparity on distance-ratio perception. Paper read to Colorado Psychological Association, Fort Collins.

WINER, B. J. (1962). *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

WOHLWILL, J. F. (1962). The perspective illusion: Perceived size and distance in fields varying in suggested depth, in children and adults. *Journal of Experimental Psychology*, **64**, 300–10.

WOHLWILL, J. F. (1963*a*). The development of "overconstancy" in space perception. In, LIPSITT, L. P., and SPIKER, C. C. (Eds.), *Advances in Child Development and Behavior*, Vol. 1. New York: Academic Press.

WOHLWILL, J. F. (1963b). Overconstancy in distance perception as a function of the texture of the stimulus field and other variables. *Perceptual Motor Skills*, **17**, 831–46.

WOHLWILL, J. F. (1964). Changes in distance judgments as a function of corrected or non-corrected practice. *Perceptual Motor Skills*, **19**, 403–13.

WOHLWILL, J. F. (1965). Texture of the stimulus field and age as variables in the perception of relative distance in photographic slides. *Journal of Experimental Child Psychology*, **2**, 163–77.

Received 25 *November* 1970

Quarterly Journal of Experimental Psychology (1971), 23, 234–242

THE EFFECT OF OVERTRAINING ON TRANSFER BETWEEN TASKS INVOLVING THE SAME STIMULUS DIMENSION

A. H. WINEFIELD AND M. A. JEEVES†

Department of Psychology, University of Adelaide, Australia

Two experiments are reported in which transfer effects between two discrimination tasks are examined following differing amounts of training on the first task. The two tasks were a conditional (successive) discrimination involving black/white cues, and a simultaneous brightness discrimination involving the same black/white cues.

In Experiment I the conditional task was presented first and in Experiment II the simultaneous task was presented first. The results of Experiment I show overall negative transfer which is not directly related to the amount of training. The results of Experiment II, however, reveal positive transfer although again the amount of training produced no significant effect. These seemingly discrepant findings are explained in terms of the difficulty of the conditional task and the development of rigid position stereotypes. It is argued that the concept of frustration-instigated behaviour is necessary to account for the results.

Introduction

In recent years the effect of overtraining on discrimination reversal has been a topic of considerable interest. Since Reid (1953) first reported data showing facilitation with overtraining, very many studies have appeared in the literature attempting to explain the overlearning reversal effect (ORE). However none of the explanations so far advanced has succeeded in accounting for the various results which have been obtained.

Mackintosh (1965) has distinguished three main categories of explanation, orienting response theories, mediating response theories, and attentional theories. He favours an attentional theory first proposed by Sutherland (1959), and later formalized by Lovejoy (1966, 1968), arguing that this approach is the most promising of those so far proposed.

The basic idea underlying the attentional approach is not a new one. It simply assumes that animals do not respond equally to all stimulus cues which are present in a learning situation, but operate selectively. This process is called attending and is held to be independent of the process whereby the animal learns to make an appropriate response. Thus discrimination learning is conceived of as involving two processes, learning to attend to the relevant cues (or stimulus dimension), and learning to make the appropriate choice response. If it is also assumed that overtraining strengthens the attending response more than it strengthens the choice response it becomes possible to explain the ORE quite neatly. Since reversal

† Present address: University of St. Andrews, Scotland.

learning involves responding to the same stimulus dimension as in learning the original discrimination, overtraining will have a facilitating effect because it protects the attending response from extinction during the early stages of reversal learning. At the same time this view is held to be consistent with the commonly observed tendency for overtrained animals to perseverate longer than controls at the beginning of reversal training. Indeed, this perseveration is claimed to demonstrate that the attending response has not extinguished.

Mackintosh (1964) has obtained strong evidence in support of the theory by showing that overtraining—as compared with criterion training—facilitates the learning of a second discimination problem involving the same stimulus dimension as the first, but interferes with the learning of a discrimination problem which involves a different stimulus dimension.

Further support for the attentional theory is provided by the work of Lawrence (1949, 1950). He demonstrated positive transfer from simultaneous to conditional (successive) discriminations and vice-versa when the same stimulus dimension was relevant in the two tasks. However, Lawrence did not investigate the effect of overtraining. The obvious relevance which Lawrence's work seems to have for an understanding of the ORE makes it desirable that his findings should be replicated and that the effects of overtraining in this situation should be examined. This was the aim of the present investigation.

Experiment I

Materials and Method

Method

Four groups of subjects were used, three experimental groups and a control. The three experimental groups were first trained to various degrees on a conditional (successive) discrimination problem in which two white stimuli or two black stimuli were presented equally often in a random order. When the white stimuli were presented subjects were required to respond on the right; when the black stimuli were presented subjects were required to respond on the left (Task C). The first group (EO) were trained to criterion, and then transferred to a simultaneous, black versus white discrimination task (Task S). The second group (E12) received 12 days of post-criterion training, and the third group (E24) received 24 days of post-criterion training, before being transferred to Task S. The control group (C) were trained on Task S only. In Task S the positions of the black and white stimuli were randomized according to Gellerman series.

Subjects

The subjects were 31 experimentally naive Wistar strain hooded male rats about 120 days old at the beginning of the experiment.

Apparatus

A discrimination box was used, consisting of a main section with a starting box and a goal box. The main section was 28 in. long, 4 in. high, and 8 in. wide. It had two aluminium top-hinged metal gates which were removable, and either could be locked by placing a nail behind it. Before each gate was a platform 5 in. wide and an air gap 3 in. wide separated the gate from the starting platform. The gates and platforms were separated by a $\frac{1}{2}$ in. wide partition. The apparatus was constructed of white painted wood and covered by a clear Plexiglass top.

Procedure

Preliminary training lasted for about two weeks During this period the subjects were deprived of food for increasing lengths of time daily up to 22 h, which was the deprivation schedule maintained throughout the running of the experiment. It took 8 days to reach this point. In addition the subjects were allowed to freely explore the apparatus for 10 min daily with the gates removed during the early stages, and unlocked grey gates in position during the later stages. When the subject seemed to have adjusted to the apparatus an attempt was made to reduce marked position preferences by means of manual guidance.

Following the preliminary training, 21 subjects began training on Task C, 7 for each experimental group. They received 10 non-correction trials per day, with 5 sec access to food following correct choices, before being returned to the home cage. Following incorrect choices they were returned to their home cages immediately. The intertrial interval was 3 to 4 min. These subjects were trained to a criterion of 90% correct over two successive days and were allocated to the three experimental groups so as to match them as closely as possible with respect to mean number of trials to criterion. On reaching criterion, Group EO were transferred immediately to Task S, Group E12 received 12 days (120 trials) of further training and Group E24 received 24 days (240 trials) of further training before being transferred to Task S. The procedural details for Task S were the same as for Task C; all subjects were trained to a criterion of 90% correct over two successive days. Group C contained 10 subjects. They began training on Task S on completion of preliminary training.

The positive stimulus in Task S was determined, for the experimental subjects, by the side preferred during the learning of Task C. A right position preference on Task C would have led to a predominance of trials where reward was received when the white stimuli were presented and no reward was received when the black stimuli were presented. In this case a right position preference might have produced a preference for white. Similarly a left position preference on Task C might have produced a preference for black. Therefore the black stimulus was made positive (on Task S) for subjects who had preferred the right side during Task C and the white stimulus was made positive for those who had preferred the left side.

For the controls, white was positive for half the subjects and black for the other half. All the subjects were housed in individual cages with water available at all times.

Results and Discussion

The mean trials to criterion are shown in Table I. It is clear that Task C was considerably more difficult than Task S. The mean number of trials to criterion on Task C for the experimental subjects was 288·93 and the mean on Task S for the control subjects was 57·40. This comparison, of course, is the one which relates to task difficulty because it is uncontaminated by transfer effects. The difference is significant at the 0·01 level ($t = 3{\cdot}57$, $df = 29$). This difference in difficulty was unanticipated particularly in view of the work of Bitterman, Tyler and Elam (1955). They found that a successive discrimination was learned more rapidly than a simultaneous discrimination under conditions where the discriminanda were spatially contiguous ("Configurational" condition). This result was reversed when the discriminanda were spatially separated ("Component" condition). In the present study the discriminanda were spatially contiguous, being separated only by a narrow partition so the big difference favouring the simultaneous learners was quite surprising.

It is also clear from Table I that the overall transfer effects from Task C to Task S were negative. Group C reached criterion in less than half the trials required

TABLE I

Mean trials to criterion: Experiment I

Group	Task C	Task S
EO	286·43	123·00
E12	303·71	195·43
E24	276·66	124·30
C		57·40

by the fastest of the experimental groups. The scores on Task S were analysed by the method of planned comparisons (Hays, 1963, Ch. 14). This method of analysis is advocated by Hays, rather than the usual analysis of variance, in cases (such as this) where the experimenter has definite questions to ask about his data to begin with (p. 475). The results of the planned comparisons are shown in Table II.

Comparison 1 (Group C against Groups EO, E12 and E24) is concerned with the overall transfer effect from Task C to Task S. Comparison 2 (Group EO against Groups E12 and E24) is concerned with the effect of overtraining on transfer, and Comparison 3 (Group E12 against E24) with the effect of different amounts of overtraining. The highly significant F ratio yielded by Comparison 1 confirms the overall interference of Task C on Task S. The non-significant F yielded by Comparison 2 is due to the comparable performance of Groups EO and E24, because Comparison 3 shows a significant difference between Groups E12 and E24.

The general findings then, are that Task C interfered with the learning of Task S, and that the effect of over-training was to first increase, and later decrease this interference, so that the amount of interference for the groups receiving most over-training was about the same as for the group receiving none.

This finding seems to contradict the results obtained by Lawrence (1949, 1950). He found positive transfer effects both from simultaneous to successive and from successive to simultaneous discriminations when the same stimulus cues were

TABLE II

Analysis of variance by planned comparisons: Experiment I

Source	*SS*	*df*	*MS*	*F*
Between groups	79,134·2	3		
Comparison:				
(1) (C against rest)	55,090·7	1	55,090·7	19·2 $P < 0{\cdot}001$
(2) (EO against E12, E24)	6,335·3	1	6,335·3	2·2 N.S.
(3) (E12 against E24)	17,708·2	1	17,708·2	6·2 $P < 0{\cdot}05$
Error (within groups)	77,303·6	27	2,863·1	
Total	156,437·8	30		

involved in the two tasks. His subjects experienced no particular difficulty with the successive task (which corresponds to our "conditional" task). The most likely reason for this is that Lawrence's procedure involved punishment for errors (falling through a trap door) and self-correction; whereas in the present experiment neither punishment nor self-correction were used.

The fact that strong negative transfer occurred is difficult to understand. Task C was chosen because it involves the same stimulus dimension as Task S, and it does not involve an incompatible choice response. On the other hand, reversal transfer necessarily involves the learning of a new, incompatible choice response from the one acquired during the original discrimination, and this, presumably, is a major source of interference. If this source of interference is eliminated it is difficult to see what other sources remain.

The point was made earlier that position preferences on Task C might have produced brightness preferences which would affect performance on Task S; to ensure that the experimental subjects did not benefit, the precaution was taken of always making the less favoured stimulus positive in Task S. It might be argued then that the experimental subjects were at a disadvantage in that they came to Task S with an initial preference for the wrong stimulus. If this were so, Task S would be like a reversal task and one might anticipate negative transfer. This argument suggests that the observed negative transfer was an artifact but it is very easily refuted by the data. Figure 1 shows the number of correct responses made by the three experimental groups during the first 6 days of training on Task S. (After 6 days subjects began to drop out, having reached criterion).

In fact the two experimental groups showing most interference, Groups E12 and E24, performed at or above chance level on the first 2 days, and on the first day one subject only in these groups chose the negative stimulus more often than the positive one. On the other hand it does appear that Group EO might have been slightly handicapped by a preference for the wrong stimulus. Four of the 7

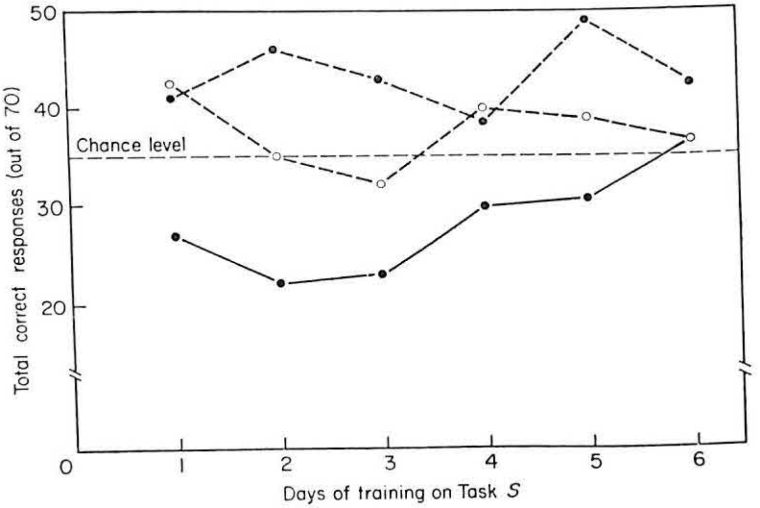

FIGURE 1. Total correct responses per group on first 6 days of Task S training: Experiment I, ●–●, Group EO; ● - - - ●, Group E24; ○ - - - ○, Group E12.

subjects in this group chose the wrong stimulus more often than the right one on each of the first 5 days of training. By the sixth day however, this group reached chance level performance, and comparing them with the controls even from this point they show no evidence of facilitation.

Perhaps a more detailed examination of the experimental subjects' performance during Task C will throw some light on their relatively poor performance on Task S. Most of the experimental subjects performed very erratically on Task C. In the majority of cases large fluctuations in performance occurred. Many subjects would repeatedly reach a high level of performance, just below criterion level, remain around that level for several days, and then drop back to chance level. Several fluctuations of this kind often occurred before the subject finally reached criterion. Behaviour during overtraining was also erratic. Having reached criterion, very few of the subjects for whom training was continued were able to maintain performance at that level. Most subjects (11 out of 14) actually dropped to chance level responding (5 or less correct in 1 day) during overtraining.

This drop in performance during overtraining was associated with an increase in position responding. In a number of cases the position preferences established during overtraining did not extinguish by the time training was terminated and were carried over into Task S. This might explain the difference in performance on Task S between Groups E12 and E24. Comparing these two groups with respect to the strength of position preferences on the last day of overtraining, the mean number of responses to the preferred side is 8·7 for Group E12 and 6·4 for Group E24. (This difference gives $P < 0{\cdot}01$ on a median test.)

Although the difference between the two over-trained groups may well be due to the development of position preference during overtraining, why these preferences should have developed at all remains unclear. It is usually assumed that post-criterion performance will not drop below the criterial level, and this appears to be true for discrimination problems of moderate difficulty (Winefield and Jeeves, 1968). No direct evidence has been reported in the case of very difficult discriminations, such as Task C, but some findings reported by Mandler (1966) seem relevant. She compared transfer effects on a brightness discrimination from a position discrimination with those from an insoluble (random) task, in both cases following different amounts of training. Interestingly, she found negative transfer from the random task but not from the position task. Even more interesting is the observation that position habits acquired during the first task persisted much longer for the random subjects than for the position subjects. This result is not inconsistent with the findings of the present study. The effect of training on a very difficult discrimination appears to be similar to the effect of training on an insoluble task insofar as both seem to lead to the development of position responding which interferes with subsequent learning.

Another possibility is that position responding per se is not a source of interference but merely a concomitant of some other factor, such as frustration or anxiety, which is the true source of interference. This explanation would fit better with Mandler's finding that position training did not interfere with brightness discrimination learning.

We have recently obtained evidence (Winefield and Jeeves, 1969) which is consistent with this latter interpretation. We attempted to induce strong position preferences in one group of subjects by means of intermittent reinforcement, and to minimize position preferences in another group by reinforcing daily position alternation. Subsequently both groups were transferred to a brightness discrimination task. In one experiment the brightness cues were absent during the position training; in another experiment, the brightness cues were present but irrelevant to the reinforcement conditions during the position training. Although both groups learned the brightness discrimination more rapidly in the first experiment than in the second, in neither experiment was there a significant difference between the groups.

These findings suggest that learning tasks which are difficult or insoluble may retard subsequent learning because of the emotional state which they produce (frustration or anxiety) and which is frequently manifested by position responding.

If this is so, we may expect that had our tasks been presented in the reverse order (the simultaneous discrimination preceding instead of following the conditional discrimination), a very different result would have been obtained. This procedure was adopted for Experiment II.

Experiment II

Method

Three groups of subjects were used, corresponding to the three experimental groups in Experiment 1. No control group was required since it is reasonable to regard the three experimental groups from Experiment 1 taken together as forming one large control group for this experiment. The procedural details were identical to those in Experiment 1 except that Task S was presented before Task C.

Subjects

The subjects were 27 experimentally naive Wistarstrain hooded male rats about 120 days old at the beginning of the experiment.

Procedure

The Apparatus was the same as for Experiment 1. There were 9 subjects in each group. Apart from this, and the interchanging of the two tasks, the procedure was the same as in Experiment I.

Results and Discussion

The mean and median trials to criterion are shown in Table III.

Comparison with the corresponding values in Table 1 reveal a strong facilitative effect from the simultaneous discrimination to the conditional discrimination. This effect is most marked in the case of the two groups which received overtraining on the simultaneous discrimination. In both cases the mean trials required to learn was reduced by more than 100. Overall the mean number of trials required to learn the conditional discrimination by the three experimental groups in Experiment I was 288·93. It is reasonable to regard the mean for these three groups as representing the number of trials required to learn the conditional discrimination independent of transfer effects from the simultaneous discrimination.

TABLE III

Mean and median trials to criterion: Experiment II

Group	Task S Mean	Task S Median	Task C Mean	Task C Median
EO	55·00	50	219·56	200
E12	50·33	48	138·56	166
E24	63·33	59	160·11	129

The results were analysed by means of a planned comparisons analysis of variance. For the purpose of the analysis the learning scores on the conditional task in Experiment I were treated as those of a control group. The analysis is given in Table IV.

As is to be expected in a difficult task there was a high degree of variability in performance on the conditional discrimination, and this is reflected in the large error variance. Nevertheless the first comparison involving the control group against the rest is highly significant. This confirms the facilitating effect of the simultaneous discrimination.

The second comparison which concerns the overall effect of overtraining is not significant on this analysis, despite the relatively large difference between the mean for Group EO and the means for Groups E12 and E24. Reference to Table III indicates that the medians are ordered differently from the means for Groups E12 and E24. This is largely due to the fact that Group E24 contained two very slow learners, so that the median may be a better index of performance than the mean. If this is so, it may be argued that a non-parametric statistical analysis would have been more appropriate than the analysis which was actually carried out. If we apply a non-parametric trend analysis to Groups EO, E12 and E24 (Jonckheere, 1954) instead of comparisons 2 and 3, we find a trend which is significant at the 0·05 level ($z = 1{\cdot}69$).

TABLE IV

Analysis of variance by planned comparison: Experiment II

Source	*SS*	*df*	*MS*	*F*
Between groups	191,077	3		
Comparison:				
(1) (C against rest)	159,378	1	159,378	6·41 $P < 0{\cdot}01$
(2) (EO against E12, E24)	29,609	1	29,609	1·19 N.S.
(3) (E12 against E24)	2,090	1	2,090	—
Error (within groups)	1,094,773	44	24,881	
Total	1,285,850	47		

In summary then, the results of Experiment II imply positive transfer from the simultaneous discrimination to the conditional discrimination and suggest that this positive transfer may increase with over-training up to 12 days. They confirm the results described by Lawrence (1949, 1950) on the acquired distinctiveness of cues. The seemingly contrary results obtained in Experiment I may reasonably be attributed to the disturbing effects of a very difficult task.

References

BITTERMAN, M. E., TYLER, D. W. and ELAM, C. B. (1955). Simultaneous and successive discrimination under identical stimulating conditions. *American Journal of Psychology*, **68**, 237–48.

HAYS, W. L. (1963). *Statistics for Psychologists.* New York: Holt Rinehart & Winston Inc.

JONCKHEERE, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, **41**, 133–45.

LAWRENCE, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, **39**, 770–84.

LAWRENCE, D. H. (1950). Acquired distinctiveness of cues. II. Selective association in a constant stimulus situation. *Journal of Experimental Psychology*, **40**, 175–88.

LOVEJOY, E. P. (1966). Analysis of the overlearning reversal effect. *Psychological Review*, **73**, 87–103.

LOVEJOY, E. P. (1968). *Attention in Discrimination Learning.* San Francisco: Holden-Day.

MACKINTOSH, N. J. (1964). Overtraining and transfer within and between dimensions in the rat. *Quarterly Journal of Experimental Psychology*, **16**, 250–6.

MACKINTOSH, N. J. (1965). Selective attention in animal discrimination learning. *Psychological Bulletin*, **64**, 124–50.

MANDLER, J. M. (1966). Behaviour changes during over-training and their effects on reversal and transfer. *Psychonomic Review Supp.*, **1**, 187–202.

REID, L. S. (1953). The development of non-continuity behaviour through continuity learning. *Journal of Experimental Psychology*, **46**, 107–12.

SUTHERLAND, N. S. (1959). Stimulus analysing mechanisms. In *Proceedings of a Symposium on the Mechanisation of Thought Processes.* Vol. II, pp. 575–609. London.

WINEFIELD, A. H. and JEEVES, M. A. (1968). Use of criterial measures in animal discrimination learning experiments. *Psychological Reports*, **23**, 1335–8.

WINEFIELD, A. H. and JEEVES, M. A. (1969). The effect of position training on brightness discrimination in the rat. *Australian Journal of Psychology*, **21**, 301–9.

Received 6 *November* 1970

Quarterly Journal of Experimental Psychology (1971) 23, 243

BOOK REVIEW

LURIA, A. R. *Traumatic Aphasia: Its Syndromes, Psychology and Treatment.* The Hague and Paris: Mouton. 1970. Pp. 479.

These intensive studies of aphasia and also of impairments in perceptual and spatial functions are based on a population of some 800 ex-servicemen, who were examined 2 to 5 months after injury, in a special neurosurgical rehabilitation hospital. They provide further evidence for the proposition that restricted speech disturbances exist and can be related to damage to different cortical areas. They indicate that site of injury and also handedness significantly affect recovery, and that closed head-injury cases show a higher percentage of recovery from aphasia than the otherwise comparable group with penetrating missile wounds. This difference in prognosis is a salutary reminder of the importance of aetiology of lesion—a factor often ignored in neuropsychological work.

The classification of symptoms and the localization hypotheses do not differ radically from those developed by earlier clinicians. However, the analysis of the main patterns of breakdown and of the techniques used for retraining is carried out within a neo-Pavlovian conceptual framework. Luria describes disturbances of acoustic and articulatory schemata associated with temporal and post-central lesions respectively, and the breakdown in sequential organization of inner speech which may follow lesions of the pre-motor area. The behavioural data and detailed case studies are of considerable interest but their interpretation remains controversial.

Luria's methods of investigating expression, comprehension, and grammatical knowledge are ingenious and productive; and he confronts the problem of distinguishing the different functional disturbances which may produce a similar clinical picture—alexia, for example. Localization is therefore approached rationally through the analysis of breakdown in the "basic components for complex speech activity." The aptness of this approach is seen in Luria's account of the different facets of speech disturbance associated with damage to different areas of left frontal cortex. This attempt to identify the components of frontal-lobe deficits, which has already proved rewarding in primate research, has barely been explored in work with neurological patients.

The study of aphasia remains a challenge, in need of physiological evidence and good post-mortem data. But experimental clinical work has made some headway and this book makes a notable contribution. It provides a valuable source of information for clinicians and research workers in the field. It might even tempt a few linguists to stray from the metaphysical splendours of competence to risk an attack on these closely-studied features of, say, semantic aphasia—difficulty with left-branching clauses, the passive tense, and the genitive, to name but a few. And those concerned with rehabilitation may have to reconsider their practice in the light of Luria's rich experience.

FREDA G. NEWCOMBE

Publications Received

ARONSON, L. R., TOBACH, E., LEHMAN, D. S. AND ROSENBLATT, J. S. (Eds). *Development and Evolution of Behavior.* San Francisco: Freeman. Pp. 656. £5.60.

BAIRD, J. C. *Psychophysical Analysis of Visual Space.* London: Pergamon Press, 1970. Pp. 321. £4.50.

BAKKER, D. J. AND SATZ, P. (Eds.). *Specific Reading Disability: advances in theory and method.* Universitaire Pers Rotterdam, 1970. Pp. 166. D.Fl.35,-.

BANNISTER, D. (Ed.). *Perspectives in Personal Construct Theory.* New York: Academic Press, 1970. Pp. 273. £4.00.

BERG, G. *Advances in Parenteral Nutrition.* Stuttgart: Georg Thieme Verlag, 1970. Pp. 243. DM38,-.

COHEN, J. *Homopsychologicus.* London: George Allen & Unwin, 1971. Pp. 192. £2.50.

DOCKRELL, W. B. (Ed.). *The Toronto Symposium 1969 on Intelligence.* London: Methuen, 1971. Pp. 267. £3.00.

FLORES D'ARCAIS, G. B. AND LEVELT, W. J. M. *Advances in Psycholinguistics.* Amsterdam: North-Holland, 1970. Pp. 454. $20.00.

GLADWIN, T. *East is a Big Bird: Navigation and Logic on Puluwat Atoll.* Oxford University Press, 1970. Pp. 241. £4.75.

GREENBAUM, S. *Verb-intensifier Collections in English.* The Hague: Mouton, 1970. Pp. 96. Dutch Guilders 14.-

HOPP, G. *Evolution der Sprache und Vernunft.* Berlin: Springer-Verlag, 1970. Pp. 167. $10.50.

HUNT, W. A. *Learning Mechanisms in Smoking.* Chicago: Aldine Publishing. Pp. 237. $8·95.

MCNEILL, D. *The Acquisition of Language: the Study of Developmental Psycholinguistics.* London: Harper & Row, 1970. Pp. 183. £3.25.

MILNER, P. M. *Physiological Psychology.* New York: Holt, Rinehart & Winston, 1970. Pp. 531.

POULTON, E. C. *Environment and Human Efficiency.* Springfield: Charles C. Thomas, 1970. Pp. 328. $15.50.

RADOM, M. *The Social Scientist in American Industry.* New Brunswick: Rutgers University Press, 1970. Pp. 210. $7.50.

SCHETTLER, G. (Ed.). *Platelets and the Vessel Wall-fibrin Deposition.* Stuttgart: Georg Thieme Verlag, 1970. Pp. 188. DM32,-.

SCHOENFELD, W. N. (Ed.). *The Theory of Reinforcement Schedules.* New York: Appleton-Century-Crofts, 1970. Pp. 316. $9.90.

CONTENTS

Printed by W Heffer & Sons Ltd Cambridge England

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Volume 23 Part 3 August 1971

PUBLISHED FOR THE EXPERIMENTAL PSYCHOLOGY SOCIETY
BY ACADEMIC PRESS LONDON AND NEW YORK

NOTES FOR AUTHORS

Only original papers are accepted for publication. All manuscripts should be submitted in *duplicate* to *Professor N. S. Sutherland, Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex, BN1 9QY, England.* One copy must be a *typewritten original* with double spacing and wide margins. Figures should be drawn to professional standards in Indian ink on tracing linen or Bristol Board; the original drawing (plus one copy) should be submitted. Allowance must be made for a reduction in size when the drawing is reproduced (the width of lines as reproduced should be between 0·25 and 0·33 mm). Explanations must be placed in the figure legend; the drawing should contain as little lettering as possible and lettering should be inserted in *blue* pencil or on a tracing paper to fit over the drawing. Tables and figure legends should be typed on separate sheets of paper. Where possible related diagrams should be grouped together to form a single figure.

Intending contributors should pay attention to the *Journal*'s conventions, and in particular to the form of headings, and sub-headings used, and references (in the Reference list titles of journals will in future be printed in full without abbreviation). The *Journal* will use the *Système Internationale* for all measurements (if terms like inches are used, metric equivalents must be quoted). The use of standard abbreviations in the text is permissible (if in doubt, write term in full the first time it is used and give abbreviation in brackets); please note that the *Journal* does not use S for subject.

An abstract must be included. Although abstracts should be as tersely written as possible, they should summarize all the main experimental findings. The length of abstracts will vary with the length and complexity of the experiments reported but should usually be between 50 and 400 words.

An initial decision on a paper will normally be reached within 2 months of receipt. Where only minor revisions are requested, date of publication will be determined by date of receipt of the paper provided a satisfactory revision is received within 1 month of the request for a revision being made. The delay between final acceptance of a paper and its publication will normally be between 3 and 6 months. The date on which the manuscript is received in its final version (apart from minor alterations—see above) is printed.

Contributors receive 25 copies of their paper free. If they require further copies they should apply to the Publishers at the time of returning proofs.

The Quarterly Journal of Experimental Psychology is published for The Experimental Psychology Society by Academic Press, London and New York. Volume 23, 1971 (4 issues): subscription rate £6.50. plus £0.40. postage inland; $16.00 plus $1.00 postage abroad. Subscription orders should be addressed to Academic Press Inc. (London) Limited, Berkeley Square House, Berkeley Square, London W1X 6BA, with the exception of those originating in the U.S.A., Canada, Central America and South America; these should be sent to Academic Press Inc., 111, Fifth Avenue, New York, New York 10003. Application to mail at second-class postage rates at Jamaica, N.Y. 11431.

Undergraduates and graduates of less than three years' standing may obtain the journal at the reduced price of £3.60. They should apply initially to the Hon. Secretary, Dr David Legge, Department of Psychology, City of London Polytechnic, Central House, Whitechapel High Street, London, E1 79F. They should not send any remittance, but enclose suitable evidence of their student status or date of graduation, such as a letter from a university teacher or copy of a graduation list.

ERRATA

A DEVELOPMENTAL STUDY OF FORM DISCRIMINATION

J. A. TAYLOR and R. J. A. WALES

Quarterly Journal of Experimental Psychology (1970) **22** (4), 720-734

p. 722, *para.* 4, line 13: insert "the figures" between "findings" and "differ".

p. 724, *Table I*: The D_{max} value of 0·37 given for the distribution of first responses to problem 4R is significant at the 0·01 level.

p. 725, *Table* 2: The total number of responses to card position 6 in problem 3 is 12, not 2.
Total correct first responses = 80.
Total responses = 288.

p. 729, *para.* 1, *line* 3: Delete "at".

p. 730, *para.* 2, *line* 12: For "98 matches" read "100 matches".

p. 730, *para.* 2, *line* 14: For "4 times" read "8 times".

p. 731, *Table IX*: Stage A, series I: delete first "Post"; the eleven subjects from 5 to 22 are all in the category "Pre".
Stage A, series I, Post: replace "21" by "22".
Stage B (i) Series III, Pre: delete one "21".

p. 733, *para* 2, *line* 14: Insert "I" after "series".

References: Birch, H. G. and Lefford, A. (1967): Visual Differentiation. . . .
Braine, L. G. (1965) Published in *Psychonomic Science*.
Garner, W. R. (1966) Published in *American Psychologist*.
Sekuler, R. W. &. Houlihan, K. Published in *Quarterly Journal of Experimental Psychology*.

THE AMOUNT SEEN IN BRIEF EXPOSURES

D. H. HOLDING

Quarterly Journal of Experimental Psychology (1971) **23** (I), 72–81

Figure 2 which appears on p. 78 should be Figure I which appears on p. 75. The captions will then be appropriate, but the explanation of symbols accompanying the existing Figure 2 should be taken to p. 75.

NOTES FOR AUTHORS

Subscription Renewal/Application

If you have renewed your subscription please disregard this notice

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Volume 24

1972

U.K.: £6.50 plus £0.75 postage and packing
Overseas: £7.50 plus £0.75 postage and packing
For Volume 24

Quotations for air mail postage provided on request

Please complete the form below and return it immediately. To avoid lapse in service and subsequent delay in obtaining missed issues when renewing your subscription *it is recommended that this application reaches us NOT LATER than December 1st, 1971.*

Should you customarily place your orders with a **subscription agency** or a **bookdealer**, please forward renewals to that agency or dealer as usual. But in that case, please do not send an advance or confirming order to us, as this may cause duplication.

ACADEMIC PRESS INC. (LONDON) LIMITED, Berkeley Square House, London W1X 6BA

Enclosed please find £........................for........renewal subscription(s) to THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, Volume 24, 1972
@ £7.25 (U.K.) or £8.25 (Overseas) inclusive of postage & packing.

Name ....
Please sign

Name ....
Please PRINT

Institution ....

Full postal address ....

....

....

....

Date

Quarterly Journal of Experimental Psychology (1971) **23**, 245–254

EEG CORRELATES OF CUMULATIVE EXPECTANCY AND SUBJECTIVE ESTIMATES OF ALERTNESS IN A VIGILANCE-TYPE TASK

ANTHONY GALE, MARY HASLUM AND VALERIE PENFOLD

Department of Psychology, University of Exeter

Occipital EEG was monitored during a slow presentation rate vigilance task. EEG samples were taken for each of 400 task events. The EEG is correlated with increases and decreases in "expectancy" built into the task. Post-trial subjective estimates of alertness parallel the EEG changes. Reaction time to "wanted signals" does not correlate with measures of pre-signal EEG.

Introduction

Work on cortical evoked potentials (e.g. Haider, 1967) and on Contingent Negative Variation (e.g. Walter *et al.*, 1964) has thrown light on the neurological substrates of expectancy in vigilance tasks. The present study examines a related problem but employs old-fashioned EEG abundance analysis in conjunction with a different experimental paradigm. Assuming that in fact the subject *does* experience shifts in expectancy and/or alertness during a vigilance task (Deese, 1955), are such shifts reflected by fluctuations in EEG? Moreover, if degree of alertness is a determinant of *speed* of detection, will reaction time to the wanted signal relate to measures of pre-signal EEG? We have constructed a simple vigilance task in which the subject's expectancy is manipulated systematically by virtue of the *types* of signal to which he is exposed. Four signal types are used, varying in terms of their power to induce cumulative expectancy. Discrete EEG samples are available for the whole duration of the task and the EEG samples associated with each signal type can be identified. After completion of the task, the subject gives retrospective estimates of experienced alertness to the different signal types. Reaction time to the wanted signal is observed. Thus, three measures of the person are taken: physiological (EEG), experiential (alertness ratings) and performance (reaction time).

Apparatus, Procedure, Subjects and Statistical Analysis

The vigilance task

Each signal consisted of 3 sequentially presented digits, followed by a plus sign. Each individual digit or plus sign was presented visually at the absolute rate of 1/5 sec. Here is a segment of the task: 2, 2, 4, +, 9, 1, 2, +, 2, 6, 8, +, 1, 3, 1, +, 1, 9, 7, +, 1, 3, 3, +, 6, 2, 8, +, 3, 2, 6, + ... The underlined sequence is the wanted signal, i.e. *three consecutive odd and unequal digits*. There were in fact four types of signal, made up as follows: (1) three even numbers (EEE +; e.g. 2, 6, 4, +; "very low expectancy"): (2) one

odd number followed by two even numbers (OEE +; e.g. 1, 2, 8, +; "low expectancy"); (3) two odd unequal numbers followed by either an odd but equal number or an even number (OOE/O = +; e.g. 1, 3, 1, + or 1, 3, 2, +; "high expectancy: no-go" and (4) the wanted signal, three odd unequal numbers (OOO +; e.g. 9, 1, 3, +; "high expectancy: go"). The principle assumption underlying the signal types is that the more odd unequal numbers there are, the higher the expectancy. There were 25 signals of each type, giving 100 signals in all for the task, each occupying 20 sec from presentation of the first digit to the presentation of the first digit of the next signal. Digits 1–9 were used and the signals were constructed with the aid of random number tables. Order of signal presentation was random, save for the constraint that no two wanted signals were adjacent. The whole stimulus sequence was stored on 8–hole punch tape. Detection, i.e. response to the third member of a wanted signal, was indicated by depression of a reaction time key.

The exceptionally slow presentation rate is determined in part by the limitations of the EEG analyzer which has a minimum epoch time of 5 sec. Moreover, for short duration EEG samples taken with *eyes open*, amplitudes are likely to be extremely small. Thus, in the absence of really powerful amplification, there is no alternative to a relatively slow presentation rate, which allows for the accumulation of higher abundance values. Presentation rate in standard vigilance tasks is typically in the region of one signal per second.

EEG recording and storage

EEG was recorded from silver/silver chloride pad electrodes, pre-soaked in 5% saline solution and held on by a rubber net. The electrodes were placed transoccipitally for bipolar recording (Cooper, Osselton and Shaw, 1969, p. 78), with the reference electrode on the non-preferred wrist. Resistance was always below 7 kilohms. The primary record was provided by a San'ei P.G. 802 Polygraph, calibrated to give 24 mm peak to trough for 100 μv with time constant at 0·3 sec. Frequency analysis was provided by a San'ei Low Frequency Analyzer EA 201 equipped with the following purpose-built pure band-pass filters: 2·0–4·5 Hz; 7·5–8·5; 8·5–9·5; 9·5–10·5; 10·5–11·5; 11·5–12·5; 12·5–13·5; 13·5–16·5 and 16·5–19·5. These filters were steep-skirted and flat-topped. Before experimental runs, filters were tested for linearity with sine wave oscillators set at the mid-point for each filter. The analyzer was set at a 5-sec epoch, coinciding exactly with the presentation of each digit in the task. The integrated outputs of the separate filters (over each epoch) were converted to digital form by a Lion Systems Development Model PD A–D Converter and punched in a 6-hole binary code on 8-hole tape. These values were subsequently printed out in decimal numbers on computer lineprint. The primary EEG record was not taken on paper but monitored continuously on a CRO.

Control apparatus

The stimulus sequence was stored on 8-hole tape (see Fig. 1) and decoded by a tape reader (A). The tape reader (A) was stepped on by the EEG analyser (B) such that digit presentation and start of analyser epoch were *simultaneous*. At the same time, EEG was recorded by the polygraph (C), displayed on the CRO (D), integrated by the EEG analyser (B), converted to digital form by the A–D converter (E) and stored on punch tape (F). Thus, integrated EEG output values were available for *each individual digit and plus sign* in the sequence, i.e. 400 values per filter per subject for the task. The stimulus was presented visually (G1) at a rate of 1/5 sec and exposed for 2 sec (controlled by the delay circuit H) on a digital display (white digit on black background 1 in. high and $\frac{5}{8}$ in. wide) set in a 24 in. square black screen. The stimulus was presented also to the Experimenter (G2) (but for the whole 5 sec epoch). He could therefore monitor the sequence throughout the task. Whenever the third member of a wanted signal was presented, a code on the tape reader (A) started a Venner millisecond timer (J). The subject stopped the timer (J) by pressing an RT key (K) held in the preferred hand. Closure of the RT key (K) also operated a code hole on the tape punch (F). Thus detections, omissions and commission errors were stored as separate events on the punch tape (F). RT was read off the timer by the Experimenter and recorded on a data sheet (25 RT's per subject, for each of the wanted signals).

Subjects

The subjects were 13 male and 7 female students, age range 18–22 years.

Procedure

When the subject had been prepared for EEG recording, he was shown written instructions which detailed the nature of the task. A printed sample signal sequence containing 7 wanted signals presented at true task density was then checked by the subject to ensure comprehension of the task. *No mention was made of the type or characteristics of the non-wanted signals,* only the wanted signal being described in detail. The subject was led to an Industrial Acoustics soundproof booth, where he sat in a barber's chair, facing the stimulus display.

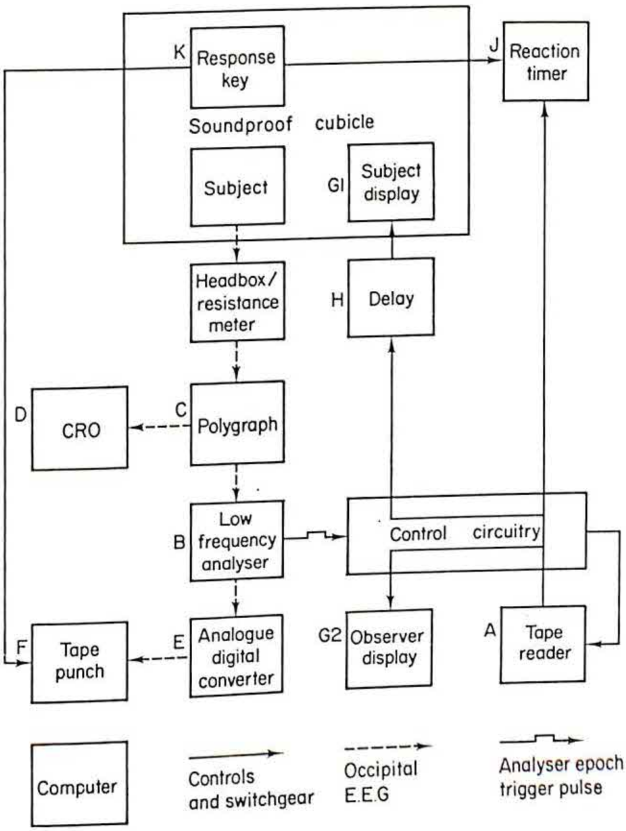

FIGURE 1. A block diagram of the experimental set-up.

The height of the chair was adjusted until the subject's line of regard was at 90° to the centre point of the display. Trial stimuli were presented at task rate and the subject trained in the use of the RT key. He was instructed to react to the third member of wanted signals as fast as possible, not to make errors, but if errors were made, to forget about them. Apart from the illumination provided by the stimulus during its two second exposure time, the subject sat in the dark. The task was then administered. Task duration was 34 min. After the task, and without prior discussion, the subject completed a questionnaire consisting of four examples of every signal type, each with different digits or plus signs members underlined (16 items to be rated). He was instructed to rate "how keyed up he felt" when the number underlined came up on the screen (e.g. 1, $\underline{3}$, 8, +). Ratings were on a five-point scale. Each subject was paid five shillings.

Data reduction and analysis

(i) *EEG*. For each subject, the nine filter values were summed and averaged for each of the four signal types, giving a total of 16 values per subject (one for each digit and plus sign, each number representing the mean of 25 trials). Thus, a value was derived for each of the 16 conditions: E, E, E, +; O, E, E, +; O, O, E/O, +; O, O, O, +. There were 8 omission errors in all (out of a possible 500). For these, the subjects mean value for detection trials was substituted. There were 20 commission errors in all (no subject with more than two errors, all occurring on signal type 3, i.e. OOE/O =). In this case actual EEG values were retained. Overall errors were so few that no analysis could be performed on them. The mean individual filter values were then collapsed for each S into three frequency bands, 2–4·5 Hz (slow), 7·5–13·5 (alpha), 13·5–19·5 (beta). For each of these three frequency bands Friedman two-way analyses of variance were computed for each signal type taken separately (i) including the plus sign value, (ii) excluding the plus sign value (Siegel, 1956). These were followed by Wilcoxon signed rank tests, comparing all columns within signal type (Siegel, 1956). Some comparisons were made between EEG output values on different signal types, comparing for example the second member of EEE+ sequence with the second member of the OOE/O sequence (see Tables I, II and Fig. 2). Since the total possible number of comparisons is large, only those which were of relevance to the hypothesis were performed. However, where differences appeared to be in the direction *contrary* to prediction, these were tested for significance.

(ii) *EEG and alertness rating*. Each subject gave 16, five-point scale ratings (one for each digit or plus sign in each of the four signal types). Again, Friedman two-way analyses of variance were computed, for the signal types taken separately followed by Wilcoxon signed rank tests (see Tables I, II and Fig. 2, where group EEG means for the alpha frequencies and subjective ratings are plotted together.)

(iii) *EEG and RT*. For each subject product moment correlations were calculated between detection RT's (i.e. RT to the third digit of a wanted signal) and (a) EEG value to the second digit of a wanted signal (i.e. value immediately prior to detection), (b) EEG value of the third digit (i.e. value covering the period for perception, detection of wanted signal and response), and (c) EEG of first digit of wanted signal minus value for second digit ("anticipation gradient"). Thus a total of 60 correlation coefficients were computed.

Results

(i) *EEG and signal type*. Mean EEG values for each signal type are given in Table I. The alpha band yields the richest results. Nevertheless, the other bands follow similar trends in most cases, but not significantly so. *EEG value goes down whenever an odd number follows another odd number (increased cumulative expectancy), and goes up when an even number follows an odd number (decreased cumulative expectancy)*. Thus, the more a sequence approximates to the wanted signal (moving left to right across Figure 2) the greater the reduction in EEG value. In addition, the "*plus*" *sign immediately following a "wanted signal" also shows increased output (decreased cumulative expectancy)*. For example, mean alpha activity under the high expectancy: go condition (OOO +) gets progressively smaller as the odd digits are presented (8·69, 8·30, 8·23) and then increases (8·94). But in OEE+ and OOE/O = +, alpha amplitude increases upon presentation of the E or the E/O =. Thus as expectancy builds up, EEG decreases; but once the subject knows that he need not prepare for response (either because he has just responded or because an E or O = appears) alpha abundance increases.

(ii) *Subjective ratings*. These are plotted against alpha activity in Figure 2. For ease of visual inspection the original scores have been subtracted from five

TABLE I

Signal type

EEG frequency/ subjective estimates	Measures and significance levels	Very low expectancy E	E	E	+	Low expectancy O	E	E	+	High expectancy "no-go" O	O	E/O=	+	High expectancy "go" O	O	O	+
	Sample stimulus	"2"	"4"	"6"	"+"	"1"	"2"	"4"	"+"	"1"	"3"	"2/3"	"+"	"3"	"9"	"7"	"+"
2–4·5 Hz "slow"	(i)	7·41	7·38	7·44	7·31	7·28	7·54	7·47	7·45	7·40	7·28	7·44	7·56	7·34	7·15	7·30	7·61
	(ii)	54·5	42·0	53·5	50·0	39·0	60·0	57·5	43·5	50·5	40·0	50·5	59·0	52·0	39·0	48·0	61·0
	(iii)		n.s.				$P < 0·05$				n.s.				n.s.		
	(iv)	43·5	32·5	44·0	—	30·0	48·5	41·5	—	44·5	33·5	42·0	—	43·0	35·0	42·0	—
	(v)		n.s.				$P < 0·05$				n.s.				n.s.		
7·5–13·5 Hz alpha	(i)	8·65	8·57	8·63	8·78	8·67	8·69	8·94	8·93	8·78	8·39	8·56	8·89	8·69	8·30	8·23	8·94
	(ii)	47·0	38·0	52·0	63·0	33·5	40·0	66·0	60·5	61·0	32·0	47·0	60·0	56·0	44·5	31·5	68·0
	(iii)		$P < 0·05$				$P < 0·001$				$P < 0·001$				$P < 0·001$		
	(iv)	41·0	34·0	45·0	—	32·0	35·0	53·0	—	51·0	29·0	40·0	—	50·0	40·5	29·5	—
	(v)		n.s.				$P < 0·01$				$P < 0·01$				$P < 0·01$		
13·5–19·5 Hz beta	(i)	5·49	5·49	5·54	5·51	5·55	5·53	5·53	5·49	5·53	5·31	5·49	5·61	5·47	5·38	5·45	5·54
	(ii)	46·5	48·0	54·5	51·0	45·0	55·5	55·5	44·0	54·5	30·5	53·0	62·0	50·0	42·0	45·5	62·5
	(iii)		n.s.				n.s.				$P < 0·01$				n.s.		
	(iv)	35·5	39·0	45·5	—	35·0	43·5	41·5	—	47·0	28·0	45·0	—	43·0	37·0	40·0	—
	(v)		n.s.				n.s.				$P < 0·001$				n.s.		
Subjective estimates	(i)	1·6	1·3	1·7	2·4	3·2	2·2	1·9	2·1	3·1	4·1	3·6	2·5	3·0	4·0	4·3	2·5
	(ii)	39·0	68·0	59·0	34·0	48·0	42·0	50·0	60·0	70·5	47·5	37·5	44·5	37·5	60·5	67·0	35·0
	(iii)		n.s.				$P < 0·001$				$P < 0·001$				$P < 0·001$		

Group means for EEG integrated output and subjective ratings: (i) means; (ii) summed ranks for (i); (iii) significance level for Friedman Analysis of variance tests on individual scores, including the plus sign; (iv) summed ranks excluding the plus sign; (v) significance levels excluding the plus sign.

(i.e. "high" alertness ratings are now low-valued and vice versa). The ratings parallel the alpha changes. For the results of Friedman analyses of variance and Wilcoxon comparison tests see Tables I and II.

(iii) *Reaction time and alpha activity*. Few of the product-moment correlations reached significance (4/60) and there was little consistency in sign. It was the subject who yielded most errors who yielded the significant correlations. The range of RT's was 148–1935 msec with a mean of 439 msec.

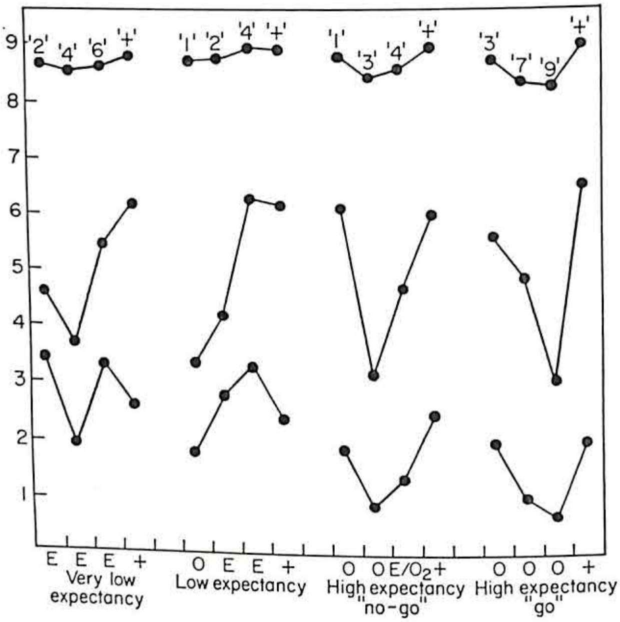

FIGURE 2. Shifts in EEG and subjective ratings for the four signal types.
Top line: EEG alpha abundance (millimetres of integrated output), each point represents the mean for 20 subjects (25 samples per subject). Sample signals are given above each point.
Middle line: summed ranks for the EEG (Scale: x10).
Bottom line: subjective ratings (the higher the score the lower the alertness). In all cases but one EEG and subjective estimates rise and fall together, point for point.

Discussion

So far as we are aware, this is the first occasion on which such discrete co-variation has been demonstrated between the EEG and an independent variable. We provide full group data to enable the reader to examine the extent of the relationship. The full power of the effect lies mainly in the alpha range. Generally speaking, even numbers and plus signs augment the EEG in relation to preceding odd numbers; odd numbers attenuate the EEG in relation to preceding odd numbers. Not only does the EEG "follow" the theoretical shifts in expectancy built into the task, but the post-trial ratings performed by the subject provide additional validation. Except for one discrepancy in signal type 1 (which we discuss below), the subject ratings and the alpha frequency EEG follow each other directionally, point for point (see Figure 2).

TABLE II

(a) "slow": 2·0–4·5 Hz

	EEE+	OEE+	OOE/O=+	OOO+
EEE+	X XX XXX			
OEE+		1 XX XXX		
OOE/O=+			5 XX X5X	
OOO+				X XX XXX

(b) alpha: 7·5–13·5 Hz

	EEE+	OEE+	OOE/O=+	OOO+
EEE+	X XX XXX			
OEE+	XXXX	X 12 11X		
OOE/O=+	X XXXX	X	1 XX 11X	
OOO+	2 XXXX	1		1 12 111

(c) beta: 13·5–19·5 Hz

	EEE+	OEE+	OOE/O=+	OOO+
EEE+	X XX XXX			
OEE+		X XX XXX		
OOE/O=+			1 X2 X15	
OOO+				X XX XXX

(d) Subjective estimates of alertness

	EEE+	OEE+	OOE/O=+	OOO+
EEE+	X XX X15			
OEE+	1 1	1 15 1XX		
OOE/O=+	1	1	1 11 511	
OOO+	1	1	XX 5	2 1X X11

Results of Wilcoxon tests of comparison both within signals and in some cases across signals. X indicates a failure to attain significance. Comparisons between odd numbers and even numbers (or plus signs) show the latter have greater EEG abundance. Comparisons between odd numbers and subsequent odd numbers show the latter to be smaller. (For the values compared, see Table I, line (i).)

The status of our retrospective ratings is questionable. It could be argued that the demand characteristics of the questionnaire *oblige* the subject to give the required ratings. Social desirability effects might induce the subject to respond in terms of how he thinks he *ought* to have felt, rather than how he *actually* felt. In addition, presentation of the signals on the questionniare is simultaneous rather than successive, as it was in the task proper. We are attempting to overcome these difficulties by providing the subject with a slider which will enable him to give ratings *during* the task.

This is of course, a relatively simple vigilance task and our use of the term "expectancy" is somewhat different from its use in the vigilance literature, where it refers to the subject's estimate of the probability of signal occurrence. However, Deese (1955) talks also of short range fluctuations in expectancy, dependent upon immediate events. This we believe, is what we are measuring here. We have provided an operational measure of the hypothetical construct.

An alternative interpretation of the data may be made in terms of "arousal" (Berlyne, 1960). The various points in the task may be ranked in terms of their "arousing properties". This would in all cases yield agreement with the expectancy predictions, with the additional advantage of distinguishing high expectancy: go from high expectancy: no go. For whereas "expectancy" does not vary for these two conditions, "arousal" does, since responding is presumably more "arousing" than not responding.

The actual absolute differences in the EEG for the 16 different conditions are in fact small. However, their *direction of change* is in accordance with our predictions and the differences are consistent and large enough to yield significance with ranking tests and non-parametric tests of comparison. We believe that the differences could be stronger (a) with different electrode placements, i.e. designed to pick up anterior-posterior changes and (b) with auditory presentation of the task with eyes continually closed. We have shown elsewhere that the Law of Initial Values holds for the EEG (Gale, Coles, Kline and Penfold, in preparation). Given the attenuation of the EEG associated with eyes open (albeit partially in the dark), there is little range left within which the EEG can respond further. An LIV effect may account for the failure of the slow and beta frequencies to yield significant results, since both start at a lower level than the alpha frequencies.

A question arises concerning the actual *nature* of the changes measured. Are they simply changes in the latency of alpha return following termination of the stimulus light? Our EEG measure is too gross to differentiate prevalence from abundance, and we are undertaking a further examination of the EEG. In the case of the final digit of wanted signals, the sample obtained contains not only EEG correlates of expectancy as such, but also those associated with confirmation of signal, decision to press, the action of pressing and presumably some appraisal of performance level.

We cannot account readily for the reduction in amplitude associated with the second digit of signal type 1. Clearly, once we have presented the subject with an initial *even* number, we have then little control over his EEG during the ensuing 15 sec. It may be the case that during this period of time, the subject is free to think about the task itself and formulate a response strategy or even to withdraw attention

from the task. In either case, one would expect EEG attenuation. However, the subjective ratings also show an increase of reported alertness at this time. Although the effect is not significant for alpha EEG (unless the value for the plus sign is included) it is contrary to our prediction that EEG values for the very low expectancy condition would fall on an horizontal line when all the subjects' scores are combined since random variations would cancel each other. As we move from left to right across Figure 2 and gain more control over the subject, the EEG and theoretical expectancy of the task yield a better fit.

The failure to obtain results for the reaction time data is disappointing. However, an increase in task difficulty might serve to improve the relationship. Certainly, the subject who found the task most difficult (as measured by number of errors) produced the only significant correlations. Few studies have shown pre-stimulus resting EEG to be a powerful predictor of reaction time. In our own case, the 5 sec EEG sample may well mask an existing short-term effect; it is certainly beyond the range of time sample normally employed in EEG/RT studies. An extension of the number of digits in the signal would provide an "anticipation gradient" with enough points for a more sophisticated statistical treatment.

Since errors were few, analysis of error data as such was not possible. We would expect false positives to be associated with extremely low EEG values and omissions to be associated with very high values. We are at present using tasks designed to generate more error data, so that such questions may be answered.

This is a pilot study which provides a basis for further work. We are extending it in a number of ways: variation in task complexity, concomitant subjective ratings (rather than retrospective ratings) which will provide individual correlations with EEG for each subject, variation in electrode placement, auditory presentation (rather than visual), the effect of individual differences, and examination of the generality of the effect on other physiological indices of attention (electrodermal activity, heart rate and respiration). An EDA study already completed confirms the EEG data (Gale *et al.*, in preparation).

Given an appropriate task and proper methods of sampling, old-fashioned (and inexpensive) EEG techniques are still capable of yielding results of psychological interest.

This work was supported by a grant No. G/169/185C from the Medical Research Council. We wish to thank Mr Arthur Burge of the Psychological Workshops, University of Exeter, for the design and construction of the control apparatus.

References

Berlyne, D. E. (1960). *Conflict, Arousal and Curiosity*. New York: McGraw-Hill.

Cooper, R., Osselton, J. W. and Shaw, J. C. (1969). *EEG Technology*. Butterworth: London.

Deese, J. (1955). Some problems in the theory of vigilance. *Psychological Review*, **62**, 359–86.

Gale, M. A., Coles, M. G. H., Kline, P. and Penfold, V. L. (1971). Extraversion-introversion, neuroticism and the EEG: Basal and response measures during habituation of the orienting response. *British Journal of Psychology*, **62**, 2.

Haider, M. (1967). 'Vigilance, attention, expectation and cortical evoked potentials'.

In SANDERS, A. F. (Ed.), Attention and Performance, *Acta Psychologica*, **27**, 246–52.

SIEGEL, S. (1956). *Non-parametric Statistics*. New York: McGraw-Hill.

WALTER, W. G., COOPER, R., ALDRIDGE, V. J., MCCALLUM, W. C. and WINTER, A. L. (1964). Contingent negative variation: an electric sign of sensory motor association and expectancy in the human brain. *Nature*, **203**, 380–4.

Received 21 *September* 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 255–260

COMPLEXITY AND COLOUR PREFERENCES IN CHICKS OF DIFFERENT AGES

JULIA C. BERRYMAN, CLARE FULLERTON AND W. SLUCKIN

Department of Psychology, University of Leicester

Five- and six-day old chicks were found to prefer a complex stimulus to a simple one, whilst 2-day old ones showed no such preference; but a red stimulus, whether simple or complex, tended to be preferred at both age levels when it was paired with a non-coloured complex stimulus. Preference for complexity may be usefully viewed in any species in the context of the behavioural development of the individual.

Introduction

Many studies indicate that adult animals will approach the more complex of two stimuli in a choice situation when other factors such as novelty are held constant (e.g. Welker, 1956; Berlyne, 1958; May, 1968). Whether this generalization is true at high levels of complexity is not precisely known, because of the difficulty of defining complexity itself. Experiments vary in the concept of complexity that they employ, e.g. size of chequer-board pattern (Simmer, 1967), randomness and the amount of contour in patterns (Karmel, 1966; Herschenson, 1964), and dimensionality of stimuli (Dawkins, 1968).

Results of developmental experiments (Karmel, 1966; Dawkins, 1968) have been less clear cut, partly due to these different concepts of complexity. Dutch (1969) however, defining complexity as "the number of elements constituting each stimulus", found that domestic fowl chicks from 10 to 14 days of age showed consistent preferences for the more complex of a series of black and white two-dimensional stimulus cards.

If preferences for complex stimuli can be thought of in terms of stimulation-seeking or exploratory behaviour then one might expect such preferences to be less marked or absent in infant animals. It would appear from imprinting studies, for example, that the most attractive stimuli for very young birds need not be complex (Sluckin, 1964; Bateson, 1966). Our purpose was to investigate the development of preferences of this kind in very young chicks and to attempt to relate them to these animals' known colour preferences. We were concerned with complexity in the sense used by Dutch (1969), but unlike him, our experiments employed only one basic kind of complex stimulus, which had either a coloured or a non-coloured ground, and a less complex stimulus lacking any pattern but containing one colour.

Experiment I

Materials and Methods

Apparatus

The rearing boxes were made of grey cardboard and measured 30 cm × 30 cm × 30 cm; each was open at the top and was heated by a 60-W bulb suspended over the centre, this bulb also providing light 24 hr a day.

The test apparatus consisted of a straight runway 76 cm long, 16 cm wide and 30 cm high; with a goal box measuring 20 cm × 20 cm × 20 cm at either end of the runway. The run was covered with movable sections of perspex to allow chicks to be placed in, or removed from, the apparatus at any point. Each goal box was lit by a 60-W bulb suspended centrally above it; thus the test apparatus was maintained at approximately the same temperature as the rearing box.

Stimulus cards were slotted in on the end wall of each goal box. The cards measured 30 cm × 8 cm and were of three kinds, as shown in Fig. 1(a), (b) and (c).

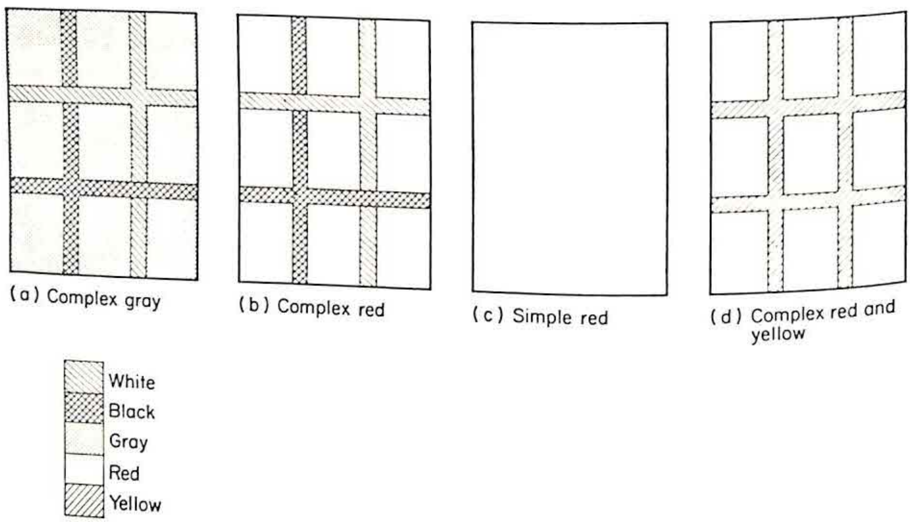

FIGURE 1.

Subjects

The subjects were 128 chicks hatched from "Cobb" eggs in the laboratory. Chicks were reared in isolation, in rearing boxes, until they were either 2 or 5 days old. Food and water were available at all times.

Procedure

Sixty-four chicks were tested in each age group. Each chick was given two choice tests. In one test the complex grey and the simple red stimuli were used, and in the other, the complex grey stimulus was again used, this time with the complex red stimulus. The chick was placed in the centre of the run, half way between the two stimuli. The test continued for 15 min, or until the chick had entered one or other goal box. If the latter occurred, the second test was begun immediately. If the chick entered neither goal box, it was recorded as a non-responder. When chicks responded, the goal box entered and the time of its entry were recorded. If the chick responded in the first test but not the second it was again classed as a non-responder.

The order of the two tests and the position of the stimuli in relation to the chicks (i.e. right and left) were varied so that all possible combinations were used an equal number of times. Thus, for each chick that responded on both tests, two choices were recorded.

Results

Four possible combinations of choices could be made in the two tests, disregarding their order. These are (a) simple red, complex red, (b) simple red, complex grey, (c) complex red, complex grey, (d) complex grey, complex grey. Tables I and II summarize the results obtained for two-day old and for five-day old chicks, respectively. Significantly more chicks responded as in (a) above in both age groups; and—as seen in Table III—significantly more two-day old chicks responded than five-day olds.

TABLE I

Responses made by 2-day old chicks
Types of response in both tests (disregarding test order)

Simple red Complex red	Simple red Complex grey	Complex red Complex grey	Complex grey Complex grey
28	3	4	6

$P < 0{\cdot}001$ (χ^2 test).

TABLE II

Responses made by 5-day old chicks
Types of response in both tests (disregarding test order)

Simple red Complex red	Simple red Complex grey	Complex red Complex grey	Complex grey Complex grey
11	0	2	1

$P < 0{\cdot}01$ (χ^2 test).

TABLE III

Responders and non-responders in both age groups

Subjects	Responders	Non-responders
2-day olds	41	23
5-day olds	14	50

$P < 0{\cdot}001$ (χ^2 test).

Experiment II

Although red was found to be highly preferred by chicks, it was not clear how this preference might be modified when complexity was introduced into a stimulus. This experiment set out simply to discover whether a simple red or a complex red stimulus would be approached more frequently by chicks in a choice test.

Materials and Methods

Apparatus

The apparatus used was that described in Experiment I, but here only the simple red and complex red stimulus cards were used.

Subjects

The subjects were 24 chicks hatched from "Cobb" eggs and reared in isolation for two days before testing.

Procedure

The testing procedure was exactly as that described for Experiment I except that only one choice test was given to each chick. Stimulus positions were reversed for half the chicks so that approach responses did not reflect directional preferences.

TABLE IV

Responses made by 2-day old chicks

	Types of response	
	Simple red	Complex red
No. of chicks	11	7

$P = 0{\cdot}24$ (Binomial test).

Results

Table IV shows that the preference scores did not differ significantly. It may be noted that the simple red stimulus contained slightly more of the highly preferred colour than did the complex red stimulus, since the latter had black and white crosses on it. Another experiment was then designed in order to equate the colour content of the two stimuli.

Experiment III

Yellow, like red, is also known to be highly attractive to chicks (Taylor, Sluckin and Hewitt, 1969). A new complex stimulus was made up for this experiment, using yellow crosses on a red ground, as in Figure 1(d).

Materials and Methods

Apparatus

The test run was as described in Experiment I. In this experiment, the simple red stimulus was used, with the complex red-and-yellow stimulus [see Fig. 1(d)].

Subjects

These were 96 chicks, hatched from "Cobb" eggs, and isolated until testing; 48 were isolated for one day and 48 for 6 days.

Procedure

Stimulus cards were slotted in, one at each end of the test runway. Chicks were tested for 15 min or until a goal box had been entered. The stimulus card approached by each chick was recorded. Stimulus positions were reversed for half the chicks in either age group.

Results

It can be seen from Tables V and VI that older chicks, although responding less, preferred the complex stimulus more than the younger chicks. The latter showed no significant preference for the complex nor for the simple stimulus.

This result in the younger animals is in accordance with that in Experiment II, which suggests that the discrepancy between the amount of colour in the two stimuli used in that experiment was probably not affecting the results.

TABLE V

Responses made by 1- and 6-day old chicks

Subjects	Types of response: Simple red	Complex red
1-day old chicks	26	16
6-day old chicks	5	14

1-day old Not sig.
6-days old $P < 0{\cdot}05$ (χ^2 test).
1 v. 6 days old $P < 0{\cdot}05$ (χ^2 test).

TABLE VI

Responders and non-responders in 1- and 6-day old chicks

Subjects	Responders	Non-responders
1-day old chicks	42	6
6-day old chicks	19	29

1-day old $P < 0{\cdot}001$ (χ^2 test).
6-days old Not sig.
1 v. 6 days old $P < 0{\cdot}001$ (χ^2 test).

Discussion

The first experiment indicated that the stimuli containing red, whether simple or complex, were both preferred over the complex stimulus lacking this colour. The attractiveness of the simple colour red can override that of a non-coloured complex stimulus, although complexity is preferred by older chicks (Dutch, 1969), when all the choice stimuli lack red.

The two subsequent experiments, however, suggest that preferences change with age. The younger animals in Experiments II and III showed no clear preference, whilst the older animals preferred the more complex stimulus, despite the decrease in the number of chicks responding. The lack of response in older chicks does give rise to some difficulties of interpretation, although the decrease is not entirely unexpected in chicks of this age since they have probably developed an attachment for their rearing pen (Taylor and Taylor, 1964), and so initial immobility tends to be prolonged in a strange environment (Sluckin, Fullerton and Guiton, 1970).

From all three experiments it may be concluded that where red is present older animals prefer complexity, whilst younger ones show no clear preference; but where red is absent in one stimulus and present in another, the latter is preferred by most subjects whether it is a simple or complex pattern.

The preference for complex stimuli in older animals has frequently been recorded (Berlyne, 1958; May, 1968), but the indication that infant animals do not show such a preference is a relatively new one. This may be a feature of early

Subjects

The subjects were 24 chicks hatched from "Cobb" eggs and reared in isolation for two days before testing.

Procedure

The testing procedure was exactly as that described for Experiment I except that only one choice test was given to each chick. Stimulus positions were reversed for half the chicks so that approach responses did not reflect directional preferences.

TABLE IV

Responses made by 2-day old chicks

	Types of response	
	Simple red	Complex red
No. of chicks	11	7

$P = 0{\cdot}24$ (Binomial test).

Results

Table IV shows that the preference scores did not differ significantly. It may be noted that the simple red stimulus contained slightly more of the highly preferred colour than did the complex red stimulus, since the latter had black and white crosses on it. Another experiment was then designed in order to equate the colour content of the two stimuli.

Experiment III

Yellow, like red, is also known to be highly attractive to chicks (Taylor, Sluckin and Hewitt, 1969). A new complex stimulus was made up for this experiment, using yellow crosses on a red ground, as in Figure 1(d).

Materials and Methods

Apparatus

The test run was as described in Experiment I. In this experiment, the simple red stimulus was used, with the complex red-and-yellow stimulus [see Fig. 1(d)].

Subjects

These were 96 chicks, hatched from "Cobb" eggs, and isolated until testing; 48 were isolated for one day and 48 for 6 days.

Procedure

Stimulus cards were slotted in, one at each end of the test runway. Chicks were tested for 15 min or until a goal box had been entered. The stimulus card approached by each chick was recorded. Stimulus positions were reversed for half the chicks in either age group.

Results

It can be seen from Tables V and VI that older chicks, although responding less, preferred the complex stimulus more than the younger chicks. The latter showed no significant preference for the complex nor for the simple stimulus.

This result in the younger animals is in accordance with that in Experiment II, which suggests that the discrepancy between the amount of colour in the two stimuli used in that experiment was probably not affecting the results.

TABLE V

Responses made by 1- and 6-day old chicks

Subjects	Types of response: Simple red	Complex red
1-day old chicks	26	16
6-day old chicks	5	14

1-day old Not sig.
6-days old $P < 0{\cdot}05$ (χ^2 test).
1 v. 6 days old $P < 0{\cdot}05$ (χ^2 test).

TABLE VI

Responders and non-responders in 1- and 6-day old chicks

Subjects	Responders	Non-responders
1-day old chicks	42	6
6-day old chicks	19	29

1-day old $P < 0{\cdot}001$ (χ^2 test).
6-days old Not sig.
1 v. 6 days old $P < 0{\cdot}001$ (χ^2 test).

Discussion

The first experiment indicated that the stimuli containing red, whether simple or complex, were both preferred over the complex stimulus lacking this colour. The attractiveness of the simple colour red can override that of a non-coloured complex stimulus, although complexity is preferred by older chicks (Dutch, 1969), when all the choice stimuli lack red.

The two subsequent experiments, however, suggest that preferences change with age. The younger animals in Experiments II and III showed no clear preference, whilst the older animals preferred the more complex stimulus, despite the decrease in the number of chicks responding. The lack of response in older chicks does give rise to some difficulties of interpretation, although the decrease is not entirely unexpected in chicks of this age since they have probably developed an attachment for their rearing pen (Taylor and Taylor, 1964), and so initial immobility tends to be prolonged in a strange environment (Sluckin, Fullerton and Guiton, 1970).

From all three experiments it may be concluded that where red is present older animals prefer complexity, whilst younger ones show no clear preference; but where red is absent in one stimulus and present in another, the latter is preferred by most subjects whether it is a simple or complex pattern.

The preference for complex stimuli in older animals has frequently been recorded (Berlyne, 1958; May, 1968), but the indication that infant animals do not show such a preference is a relatively new one. This may be a feature of early

stages of perceptual development (Fantz, 1958) or else the absence of any marked exploratory tendency in very young individuals. Obviously, more detailed examination of this phenomenon is required before any useful generalizations can be made.

We should like to thank Richard Parsons, Stuart Bailey and Patrick Kerrigan, for their technical assistance, and are grateful to John Dutch for his helpful advice.

References

BATESON, P. P. G. (1966). The characteristics and context of imprinting. *Biological Reviews*, **41**, 177–220.

BERLYNE, D. E. (1958). The influence of the albedo and complexity of stimuli on visual fixation in the human infant. *British Journal of Psychology*, **49**, 315–8.

DAWKINS, R. (1968). The ontogeny of a pecking preference in domestic chicks. *Zeitschrift für Tierpsychologie*, **25**, 170–86.

DUTCH, J. (1969). Visual complexity and stimulus pacing in chicks. *Quarterly Journal of Experimental Psychology*, **21**, 63–6.

FANTZ, R. L. (1958). Pattern vision in young infants. *Psychological Record*, **8**, 43–7.

HERSCHENSON, M. (1964). Visual discrimination in the human newborn. *Journal of Comparative and Physiological Psychology*, **58**, 270–6.

KARMEL, B. Z. (1966). Randomness, complexity, and visual preference behaviour in the hooded rat and domestic chick. *Journal of Comparative and Physiological Psychology*, **61**, 487–9.

MAY, R. B. (1968). Pretest exposure, changes in pattern complexity and choice. *Journal of Comparative and Physiological Psychology*, **66**, 139–43.

SIMNER, M. L. (1967). Age changes in preference for visual patterned stimulation in the newly hatched chick. *Psychological Abstracts*, **41**, 1343.

SLUCKIN, W. (1964). *Imprinting and Early Learning*. London: Methuen.

SLUCKIN, W., FULLERTON, C. and GUITON, P. (1970). The development of aversive responses in domestic chicks. *Revue du Comportement Animal*, **4**, 73–9.

TAYLOR, A., SLUCKIN, W. and HEWITT, R. (1969). Changing colour preferences of chicks. *Animal Behaviour*, **17**, 3–8.

TAYLOR, A. and TAYLOR, K. F. (1964). Imprinting to a static feature of the home environment. *Nature*, **204**, 1117–8.

WELKER, W. I. (1956). Some determinants of play and exploration in chimpanzees. *Journal of Comparative and Physiological Psychology*, **49**, 84–9.

Received 23 October 1970

Quarterly Journal of Experimental Psychology (1971) 23, 261–267

EVIDENCE FOR A TIME CONSTANT IN RATE OF SPEECH UNDER DELAYED AUDITORY FEEDBACK

D. SALTER AND J. K. CLARKSON

Department of Psychology, University of Sheffield

This experiment investigates an apparent discrepancy in experimental measurements of the effect of texture predictability upon reading disruption under delayed auditory feedback (DAF). By measuring relative DAF decrement in three different ways, it is shown that the previous findings can be related; Fillenbaum's hypothesis of increased disruption by DAF with an increase in predictability of the material is rejected, less disruption being obtained after practice on a particular passage. Almost identical ratios of DAF rate divided by normal rate are found irrespective of the type of reading material and stage in practice. This has not been reported previously and suggests that behaviour under DAF may be related to behaviour under normal conditions by a multiplicative constant. These results are also consistent with the notion of limited channel capacity and the partitioning of attention between two sources of information.

Introduction

A number of studies using delayed auditory feedback (DAF) have reported on the amount of *relative* speech disruption (i.e. normal to DAF conditions) with different reading materials. Their findings are not consistent.

Fillenbaum and Wiesen (1961), for example, concluded that low word-order approximations appear to be *less* affected by DAF than continuous prose. Their report was based on calculations derived from earlier work carried out by Elliott (1956). Two subsequent experiments (Fillenbaum, 1963) with different materials appeared to corroborate the earlier observation. These results conflict with those of a study by Treisman (1965) in which low word-order approximations appear to be *more* disrupted by DAF. At the conclusion of her report Treisman noted the discrepancy between her results and those of Fillenbaum and Wiesen (1961) but was unable to explain it. The present study offers evidence which may resolve the difficulty.

In these studies relative disruption of reading has been measured in a variety of ways: (1) by a ratio of the number of correct-letters per second (Fillenbaum and Wiesen, 1961), (2) by a difference in the times taken to read the passage (Fillenbaum, 1963); and (3) by the time per word as a function of information in bits (Treisman, 1965). In some studies subjects have acted as their own controls; in others separate groups have been used. Discrepancies between results could therefore be due to differences in the measure used or in experimental method.

To explain his findings, Fillenbaum (1963) proposed that "The more automatized a speech task, the more exclusively it may come to depend upon concurrent

feedback for its continuation and proper execution, and therefore the more it should suffer in execution with tampering in such feedback, under conditions of DAF" (Fillenbaum, 1963). As a direct test of this explanation, the following experiment included repeated readings of the same material under DAF. If the hypothesis is correct, then increased disruption from DAF is predicted for the later readings.

Materials and Methods

Stimulus material

A prose passage, 260 syllables (200 words) in length, was chosen from G. Orwell's essay "Down the Mine". The passage was presented in two versions: (1) with words in the original order (PR); (2) with words in random order (RW), but with punctuation retained at its former locations. This method of constructing the stimulus material has the advantage over previous studies using word-order approximations that the same words are common to both conditions.

Design

Each subject read each passage three times. Initially both passages were read with DAF, then without DAF or headphones, and finally for a second time with DAF. The order of prose and random word passages over the three readings was constant for any one subject, but was reversed for alternate subjects to counterbalance any order effect. A warm-up task of reading under normal conditions with separate pasages was given to reduce the possibility of differential reading improvement with the two types of material.

Subjects

Twenty graduates and undergraduates of the Department of Psychology, University of Sheffield were tested (mean age 21 years 7 months). All had previous experience with DAF and none suffered from any hearing impairment.

Procedure

A headrest adjusted for each subject ensured that a constant distance was maintained from the microphone at all times. For the warm-up task, subjects read a passage of random words, without DAF, at the fastest rate of speech they could manage. They were then asked to read aloud a sheet of instructions telling them:

(1) to read all passages at an optimal rate for accuracy and speed, that is "at a rate quicker than normal" but one at which they would not trip over themselves and make mistakes;

(2) to try not to omit words;

(3) in the event of saying the wrong word, to go on to the next rather than go back to repeat the whole phrase;

(4) in the case of mispronunciation,to go on and finish the word.

Attention was called to these instructions at the beginning of each block.

A delay interval of 225 msec was used throughout for DAF conditions. The sound pressure level of the subject's speech at the headphones was approximately 90–95 dB (sufficient to mask bone conduction). A few seconds elapsed between blocks while headphones were removed or fitted by the experimenter.

Results

The raw data are summarized in Table I. The syllable was selected as a convenient unit for analysis. It is less variable in length than the word, and more appropriate for an articulation score than correct letters per second.

TABLE I

Means of raw data for reading times and rates (conditions in order of presentation)

	DAF 1		NORMAL		DAF 2	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
(a) Times (sec)						
Prose	65·3	12·7	46·8	3·5	59·3	10·9
Random	91·8	17·6	67·3	7·8	85·1	14·2
(b) Rates of syll/sec						
Prose	4·105	0·684	5·592	0·433	4·502	0·674
Random	2·923	0·506	3·898	0·469	3·137	0·533

From these data three measures of performance decrement were derived for prose and random words over session one and session two:

(1) times: DAF minus normal;

(2) rates: normal minus DAF;

(3) ratios: DAF rate divided by normal rate.

Table II gives analysis of variance results for (1), (2) and (3), and Figure 1 illustrates prose v. random word differences over sessions for each measure.

TABLE II

Results of three analyses of variance for measures of relative disruption when reading prose, and the same words in random order, under conditions of normal and delayed auditory feedback

	(1) Time differences		(2) Rate differences		(3) Ratios	
	F	P	F	P	F	P
PR v. RW	13·85 RW > PR	$<0·01$	35·22 PR > RW	$<0·001$	0·62	n.s.
Sessions	24·13	$<0·001$	50·35	$<0·001$	39·40	$<0·001$
Presentation order	0·29	n.s.	0·03	n.s.	0·03	n.s.

F ratios with 1, 18 degrees of freedom throughout. The only significant interaction found was with measure (2): PR-RW × sessions ($F = 7·25$, $P < 0·025$), which is illustrated by the convergence of the two lines in Figure 1(b).

Order of presentation of the type of passage is not significant, and there is no significant interaction between type of passage and order. Hence, there is no evidence for practice transfer from a passage of prose to the same words in random order, or vice versa. Testing between means appropriate to the design (Winer, 1962) was carried out. The time difference measure and the rate difference measure are both highly significant for PR and RW (time: first session $P < 0·01$, second session $P < 0·02$; rates: first and second session $P < 0·001$). But the differences for times and rates are in *opposite* directions. As can be seen from Figure 1, it could be concluded from time differences that there is significantly greater disruption under the DAF condition with an increase of information content. This result is consistent with the data of Treisman (1965). However, for the rate

differences the disparity between DAF and normal decreases with an increase in information content; there is less disruption under DAF with increased information as Fillenbaum (1963) has claimed. Again the result is highly significant. However, it is of much more interest that by the ratio measure neither PR nor RW is relatively more disrupted with DAF. All three measures concur in showing

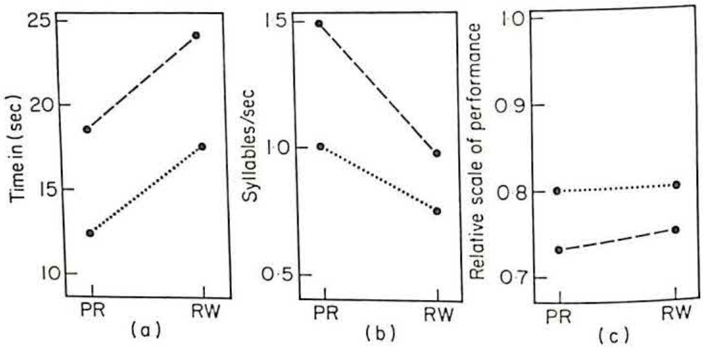

FIGURE 1. Three measures of relative performance with passages of prose (PR) and the same words in random order (RW) in two sessions with DAF.
(a) Time differences (DAF time − NORMAL time).
(b) Rate differences (NORMAL rate − DAF rate).
(c) Ratios (DAF rate/NORMAL rate).
– – –, First session; . . . ., second session.

marked improvement over sessions for both prose and random words. Relative disruption decreases for PR (measures: (1) $P < 0{\cdot}01$, (2) $P < 0{\cdot}001$, (3) $P < 0{\cdot}001$), and for RW (measures: (1) $P < 0{\cdot}002$, (2) $P < 0{\cdot}002$, (3) $P < 0{\cdot}002$) from the first to the second session with DAF.

Discussion

The hypothesis of Fillenbaum (1963), stated in the introduction, predicts *more* disruption in a second session with DAF where the PR or RW passage is being read for a third time. Evidence based on all three measures gives grounds for decisively rejecting this explanation, since the high significance levels obtained for practice give strong support for the contradictory hypothesis.

It is easily demonstrated that mathematical transformations of data can eliminate a significant interaction in variance analysis, but it is rare in practice to find a situation where such a transformation will produce a further, clearly significant interaction component but in the opposite direction. Where time measures have been used to show increasing DAF to normal differences with increasing information, as in Treisman's (1965) experiment, transformation to rate measures might be expected to reveal a very different picture. The validity or otherwise of time measures is not in question here—they may indeed serve as a better basis for theorizing in this field. There is, however, the possibility as Fairbanks (1955) and Carroll (1966) have pointed out that duration measures may be positively skewed, and for this reason a geometric mean is sometimes preferred when analysing time data. The analysis of different measures from the same data, illustrated in Figure 1, shows

clearly the source of contradiction between the results of Treisman and Fillenbaum and provides the explanation sought by Treisman in 1965.

The very small value of F for the ratio measure in Table II, together with the differences obtained by time and rate measures, suggest the following hypothesis. Given the same material, regardless of the complexity of its information content, time taken under conditions of DAF is equal to $1/k$ times the time taken under normal conditions, where k is a constant less than one. This simple formula has the property that where a time difference measure is used, measurement will increase with increased information. For example: with a passage of prose, if the normal time equals t and DAF time equals $(1/k)t$, DAF time minus normal time will then be equal to $t(1-k)/k$. With a passage of random words, normal reading time increases to $t + x$; DAF time increases to $1/k(t + x)$. The difference measure is given by $(t + x)(1 - k)/k$. Hence a comparison between PR and RW will reflect the increase in normal times with RW. Similarly, it can be shown that a rate measure will show a decrease in difference between DAF and normal with increasing information. With a ratio measure, PR and RW yield non-significant differences as only the multiplicative constant k is compared.

It is possible that this hypothesized constant k is only likely to emerge clearly where certain experimental controls are used. There are two reasons why k values constructed from Treisman's data do not exhibit a similar degree of constancy. Different subjects were tested in the DAF and normal conditions. Second, for each group the order of passages was randomized so that any practice effect in the DAF group would add to the variance of the measure taken. (It can be seen from Treisman's Figure 1 that the means for DAF appear more variable than those for other conditions.)

It could be suggested that Treisman's results differ from the present ones in that the intercept also changes with DAF so that the ratio of DAF to normal times was not constant in her experiment. We would point out, however, that the intercept of the time axis is not experimentally determined but occurs merely as the extrapolation of a regression line to a point where information is zero. Such a hypothetical intercept has little relevance to ratios constructed from the data.

In Fillenbaum and Wiesen's (1961) article ratios for third and fifth word-order approximation, and for prose, show the same order of constancy as in the present experiment. The significant reduction in decrement they observe for zero and first word-order approximations may be traced, we suggest, to their use of *correct* letter rate per sec. This is likely to include a bias for qualitatively different stimulus material.

One deduction that might be made about the constant k is that it could be affected by learning but that this learning, resulting in an increase in k, would be equal for both PR and RW conditions. (If k increases, the reading time under DAF is reduced.) The observed "k" values, namely the DAF-normal ratios for each session fit this pattern remarkably well. From Figure 1(c), it can be seen that the two lines are very nearly horizontal, and thus parallel. Estimates of k for prose and random words are respectively: PR 0·732 (0·094) and RW 0·751 (0·103) for the first session and PR 0·803 (0·091) and RW 0·805 (0·100) for the second session. It is a necessary implication that if NORMAL/DAF is a constant for both prose

and random words then PR/RW will be a constant over normal and DAF conditions. That this holds within very small limits can be seen from the rates given in Table I. The PR-RW ratios are 0·712 for DAF1 and 0·697 for NORMAL, and 0·697 for DAF2. As might be expected from estimates of k, correspondence between DAF ratios and normal ratios increases with practice.

The theoretical implications of k, whilst not dissimilar from those originally formulated by Treisman, may perhaps be given a more concise statement. For a given time interval t, for normal reading, a constant t' is added under DAF conditions which is independent of the number of words occurring during t. It might be expected that information processed during t remains constant whether prose or random words are uttered, and t' also remains constant. An analogy might be drawn between a fast and a slow computer; the greater the number of bits to be processed the greater the disparity in total time taken by each.

Treisman (1965) proposed two ways in which subjects monitor the different passages: either at constant intervals for variable times, or at variable intervals for a constant time. She presents some evidence to support the latter alternative. The results given here extend that finding by indicating that the division of processing capacity between the primary information source (reading material) and the secondary information source (DAF) remains constant for PR and RW. (Treisman's evidence is not inconsistent with the simple view that monitoring rate may remain constant in a given time t.) This is due to the fact that the subject varies his rate of reading with the amount of information of the material (see Table Ib). With repeated readings of the same passage, capacity made available through practice might either be given over to increased monitoring or allotted to processing the main reading task. The increased reading rate over sessions and the higher ratio (see Figure 1) indicate that reduced processing of DAF occurs. A similar type of finding is that of Herman (1965). He notes that when subjects perform simultaneous tasks (auditory tracking, and discrimination) they increase information transmission on the tracking task and reduce it on the discrimination task with practice although the total amount of information transmitted remains constant. His evidence suggests that channel capacity may be partitioned in different ways between simultaneous tasks, but that the total amount of information is primarily the function of the information capacity of the channel rather than the information characteristics of the task.

We submit that similar conclusions are to be drawn from the present evidence on reading aloud with DAF.

One of us (D. S.) is indebted to the Social Science Research Council whose Studentship Award made this work possible.

References

Carroll, J. B. (1966). Problems of measuring speech rate. Presentation to Conference on Speech Compression, University of Louisville, Kentucky.

Elliott, C. R. (1956). Disruptions in reading to delayed auditory feedback as a function of contextual constraint. Unpublished Ph.D. dissertation, University of North Carolina.

FAIRBANKS, G. (1955). Selective vocal effects of delayed auditory feedback. *Journal of Speech and Hearing Disorders*, **20**, 333–45.

FILLENBAUM, S. (1963). Impairment in performance with delayed auditory feedback as related to task characteristics. *Journal of Verbal Learning and Verbal Behaviour*, **2**, 136–41.

FILLENBAUM, S. and WIESEN, R. (1961). Contextual constraints and disruptions in reading with delayed auditory feedback. *Journal of the Acoustical Society of America*, **33**, 1800.

HERMAN, L. M. (1965). Study of the single channel hypothesis and input regulation within a continuous, simultaneous task situation. *Quarterly Journal of Experimental Psychology*, **17**, 37–46.

TREISMAN, A. M. (1965). Reading rate, word information and auditory monitoring of speech, *Nature*, **205**. 1297–300.

WINER, B. J. (1962). *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Received 2 November 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 268–274

CODING FACTORS IN TRANSCRIPTION

L. H. SHAFFER AND ANN FRENCH
Department of Psychology, University of Exeter

Two experiments looked at aspects of stimulus grouping by trainee typists and tested for the possibility of response grouping. Manipulating text preview indicated that typing speed declined if the next word was not wholly on display but, conversely, there was little gain if it was displayed much in advance of its being typed. Some of the effect of practice was specific to the words actually typed: words not presented during practice were typed slower than practised ones; others given reinforced practice were typed quicker than the rest. An analysis of response latencies provides evidence that response grouping sometimes occurred.

The purpose of this paper is to examine some aspects of the notion of grouping in transcription skill, in the particular context of typing. By grouping we mean a coding in transmission based upon sets of elements rather than single elements. The concept was used by Bryan and Harter (1899) to account for the step-like changes of performance with practice in morse reception. It was taken up by Craik (1948) and by Lashley (1951) to explain respectively the speed and fluency achieved in various transcription skills.

A simple model to explain the temporal characteristics of transcription, such as typing, reading aloud and playing music, is as follows: an input (perceptual) system translates stimuli into response codes and stores them in a buffer (primary) memory; an output (response organizing) system accepts the codes and converts them into motor output. The buffer is postulated in case the input system can recycle and accept new text while a queue of codes await conversion by the output system (cf. Shaffer, 1971). Within this model one must ask whether grouping can occur at input, at output, or both, and in each case what are the organizing principles of grouping.

The temporal aspects of typing have recently been the subject of intensive research (Hershman and Hillix, 1965; Shaffer and Hardwick, 1968, 1970) and it seems sufficient, in accounting for the results, to suppose that grouping occurs at input alone. In the fine detail of response there may be reason to believe that successive finger movements are not independent: if a transition may involve wrist movement then there can be no invariant response to execute a particular key press. Thus response commands may have to be context-sensitive rather than context-free (cf. Wickelgren, 1969). However, this Markovian property of responding in typing does not require a postulate of response grouping as might, for instance, the expressive and rhythmic features of phrasing in speech and playing music. Also, errors such as letter order inversions in typing and spoonerisms in speech, which Lashley cited as evidence for response grouping, can as easily be considered faults in buffer storage, since they are found in the study of primary memory.

The idea of response grouping in typing is that for familiar letter sequences (ing, ion, . . .), words (and, the, . . .) or phrases (Dear Sir, . . .) the typist has learned to produce arpeggios of fast responses. A good reason for resisting this idea is that it predicates large amounts of response learning, particularly when it is extended as a general principle of fast typing. Leonard and Newman (1964) trained subjects to type a very restricted vocabulary. There was strong evidence of grouping, and with only ten different words subjects may have been encouraged to learn response groups, but the analysis failed to distinguish whether it occurred at input or output. Thomas and Jones (1970) asked professional typists to type prose ignoring all spaces, and the text was such that natural spaces were suppressed in the prose and new ones put in at arbitrary regular intervals. They found an asynchrony between pauses in output and the boundaries in text units. It may be that the experimental conditions induced response clustering, but this is different from a constructive principle of grouping. Of the experiments that follow some look at aspects of input grouping and others for the possible emergence of response grouping by subjects learning to type.

Experiment I

The aim was to take subjects with little previous typing experience and train them to type on a special keyboard, up to some reasonable level of proficiency, and then give them a series of texts to examine the grouping tactics they had developed.

One test gave subjects texts containing words they had not met during training. If they had formed response groups on words then they should be slower on such texts than on familiar ones. Other tests examined the extent of use of preview.

Method

Apparatus

The keyboard had 12 letter keys and a space bar arranged as follows:

S I L   C O N
R A D   H E T

(space)

It had automatic control of display renewal and data recording via transistorized circuits. The display was projected by a bank of six inline units and a shift register transposed the symbols from right to left, by one symbol, immediately a key was pressed. The number of symbols actually on display in any trial could be reduced by use of shutters. Since the subject had always to type the leftmost symbol the others on display were available for preview. The apparatus provided a complete record of all response latencies and errors. More details of the apparatus can be found in Shaffer and Hardwick (1970).

The alphabet A C D E H I L N O R S T contains the most commonly occurring letters in English and provided a vocabulary of over 1000 real words 1–6 letters long. A subset of these words, to be described later, was kept aside and the rest were used to make up 36 texts, all containing 7 words of each word length.

Design

Fourteen student subjects were given the same treatment. They had three practice sessions lasting 1 hr each, in which texts were exposed one letter at a time and they were encouraged to learn to touch-type. There followed eight practice sessions in which text

exposure was 6 letters and in each session there were 18 texts taken from the 36. Subjects were instructed to type these as fast as they could with few errors.

Following practice there were two test sessions:

Test 1 contained 14 trials in which 7 conditions were presented twice each in scrambled order. Five of the conditions used familiar texts and different levels of text exposure, 1, 2, 3, 4 or 6 symbols. A sixth condition used new texts in which half the 3 letter words and all the longer ones appeared for the first time in the experiment. These words had been randomly selected from the vocabulary. In the seventh condition the longer words were nonsense, created by arbitrarily joining syllables from different words in the vocabulary. All texts had the same distribution of word lengths. Test 2 presented 21 texts and in any text all (real) words had the same length. These texts were presented at different levels of exposure to give the conditions indicated in Table I by crosses.

TABLE I

Experimental conditions used in Test 2. Experiment I

	Level of Exposure					
	1	2	3	4	5	6
Word length 1	x	x	x	x	x	x
2		x	x	x	x	x
3			x	x	x	x
4				x	x	x
5					x	x
6						x

Results

Here and subsequently the data presented will only be of inter-response times (IRTs). There is a record of errors and they give a correlative picture: if a text or display condition led to longer IRT it also led to more errors (cf. Shaffer and Hardwick, 1970).

Test 1

An analysis of variance was carried out on the 5 conditions in which text exposure varied. The mean IRT in each condition is shown in Table II. Ignoring the two rightmost entries IRT decreased with increasing exposure, at a rate that changed at 3 symbols. The exposure effect is significant, $F\ (4,52) = 13{\cdot}2$, $P < 0{\cdot}001$, but

TABLE II

Mean IRT (msec) in the conditions of Test 1. Experiment I

Exposure						
1	2	3	4	6	6N	6S
754	622	523	493	463	498	552

none of the adjacent pairs of conditions differ significantly on Scheffé S tests: starting from the left the first contrast significant at the 5% level is between 1 and 3 symbols, and starting from the right it is between 6 and 2 symbols.

The three conditions at the right of Table II are, from left to right, of familiar, new-word and nonsense texts and IRTs increased in this order. The differences were significant in an analysis of variance, $F(2,26) = 61$, $P < 0{\cdot}001$, and the

pairwise differences were significant on S tests with $P < 0.01$. Further brea down between old and new texts showed that there were differences at all wo lengths between old and new words.

Test 2

The main result of interest in this test is given in Table III. There was effect, diminishing with word length, that if a word was in full view one or tw symbols before it was typed then it was typed faster. Additional preview h little effect. Put another way, it is as though a new word was read only wh typing the previous word was near completion.

TABLE III

Mean IRT (msec) in the conditions of Test 2. Experiment I

		Exposure					
		6	5	4	3	2	1
Word length	1	392	384	387	400	456	549
	2	386	385	386	403	459	
	3	433	440	451	459		
	4	475	462	510			
	5	497	496				
	6	482					

Discussion

It should be borne in mind that the results obtained here describe an arbitrar stage in practice. Their value is in showing how one may test hypotheses abou the tactics of grouping in typing—or any form of transcription.

The effect of text preview replicates the earlier one found by Shaffer and Hard wick (1970), but a new interpolated level shows that gain in speed with previe decelerated beyond 3 symbols. As exposure increases longer words come int full display and the data on preview from Tests 1 and 2 taken together indicate tha speed approached asymptote when the next word to be typed was in full display Since there was little further gain from additional preview they also suggest tha input grouping seldom went across words.

Test 1 also shows that practice in typing was partly specific to the words actuall typed. If this had not been the case it would have refuted the hypothesis tha there is response grouping on words. As it is the result is ambiguous—it coul indicate response grouping or a refinement of input grouping. The other part o the result shows that new words were nevertheless more privileged in transcriptio than nonsense words having syllabic structure. This part of the experimen extends the result of Leonard and Newman (1964) since it is based on a muc larger vocabularly.

Experiment II

An alternative tactic, to discover if response grouping is used, is to reinforc selectively certain words by repetition and this was done here. There was also stringent test for response grouping based on the stochastic latency model outline earlier.

In the model let the input system translate stimuli into response codes with a random latency τ and the output system convert codes into responses with a random latency t. As long as there is a queue of codes in memory, responses can be made at a rate $1/_t$; as soon as the queue becomes empty the latency of the next response will be $(\tau - \lambda) + t$, where λ is the temporal overlap between initiating a new translation of stimuli and converting a residual queue into response.

If there is input grouping it is assumed that for groups of size k, $\tau_1 < \tau_k < k\tau_1$, where τ_k is the latency of translating the group. Similarly if there is output grouping it is assumed that for groups of size m the first response has a latency $t_m > t_1$ and the subsequent responses have latencies $t_r < t_1$ such that $t_m + (m-1) t_r < mt_1$. Thus we motivate the notion of grouping by supposing that it permits faster overall responding.

The model predicts that the distribution of response latencies, or IRTs, in a trial should be a mixture of latencies $T = t$ and $T = (\tau - \lambda) + t$ (inserting appropriate suffixes for assumptions about grouping). One finds that actual distributions typically have a marked peak and a long right tail (Shaffer and Hardwick, 1970) and the variation in different conditions makes it reasonable to identify the peak, with the latencies $T = t$.

The test for response grouping can now be stated: response grouping on a word occurs if all its letters after the first are typed with latencies falling in the first quartile of the latency distribution of that subject. The choice of the first quartile as cut-off point is arbitrary but it provides a reasonably severe criterion. It defines a region within the distribution of latencies associated with t, since at the stage of practice tested here the peak is typically the major part of the distribution.

Method

Ten subjects who had served in the first experiment were invited to come back for 5 more test sessions. In each session there were 18 trials: in each trial text exposure was 6 symbols and a text was made up of words all the same length; word lengths could be 1 to 6 letters and all lengths appeared three times in the session. Each session used the same set of texts in varying order. In each text there were two words that recurred in random locations and these comprised about a quarter of the text, while none of the other words were repeated within the text. In their respective texts the words were I and A, IS and DO, AND and THE, ALSO and DEAR, HELLO and SINCE, ARTIST and STREET.

Results

It is sufficient to look at the latency data in the last session. Table IV shows mean IRT for recurrent and non-recurrent words in texts of different word length. An analysis of variance shows that the effect of recurrence was significant, $F(1,9) = 71$, $P < 0{\cdot}001$. The effect of word length was significant, $F(5,45) = 8{\cdot}9$, $P < 0{\cdot}001$, and the table shows a general tendency for IRT to increase with word length. There was also a significant interaction between these two variables, $F(5,45) = 9{\cdot}5$, $P < 0{\cdot}001$, and it is seen that the difference in IRTs between recurrent and non-recurrent words was not uniform with word length but was minimum in 2 letter words and maximum in 5 letter words. The origin of this form of interaction is rather obscure.

The test for response grouping described above required a search within the

TABLE IV

Mean IRT (msec) in recurrent and non-recurrent words in texts of different word length. Session 5, Experiment II

	Word length					
	1	2	3	4	5	6
Recurrent	302	315	293	327	300	356
Non-recurrent	366	341	383	421	431	434

trial data of each subject, in trials with words longer than 1 letter, for the event specified, namely that IRTs to all letters after the first in a word should be in the first quartile of the distribution. For convenience we call this event *Q*.

A count of the occurrence of *Q* events was recorded for each subject and the aggregate picture for ten subjects is shown in Table V. Allowing for word

TABLE V

Average frequencies per subject of words classified as response grouped (Q *events*)*; where* n *is word frequency,* E(Q) *is the expected number of* Q *events occurring by chance and* G(Q) *the number that would be exceeded by chance with probability* P < 0·05. *Session* 5, *Experiment II*

	Word length	Observed frequency	*n*	Lower bound *E(Q)*	Lower bound *G(Q)*	Upper bound *E(Q)*	Upper bound *G(Q)*
Recurrent	6	0·1	24	0	0	0	1
	5	1·0	30	0	1	0	1
	4	4·4	30	0·5	2	1	4
	3	7·0	36	2	6	5	8
	2	19·8	42	10·5	16	21	26
Non-recurrent	6	0·3	72	0	1	0	2
	5	0·4	78	0·5	2	1	2
	4	2·7	90	1·5	4	3	6
	3	5·5	99	6	11	14	20
	2	45·9	123	31	39	61	70

frequency, *Q* events occurred more often among recurrent than non-recurrent words, except in 6 letter words. In order to evaluate the results two models provide respectively what seem to be reasonable lower and upper bounds of chance occurrence. The first assumes that all responses were equally likely to be fast ones; the second assumes that fast responses were not made to the first letters of words but were equally likely to appear elsewhere. In each case *E*(*Q*) is the expected frequency and *G*(*Q*) the frequency that would be exceeded with probability $P < 0{\cdot}05$.

In Table V the frequency of *Q* events in recurrent words was always above the expected score, on either chance model. In 4 letter words it was statistically significant on either model, while for 2 and 3 letter words it was significant only on the lower bound model. In non-recurrent words the frequencies were usually bounded by the expected scores on each model.

The result may not seem compellingly to favour the hypothesis of response grouping. However the analysis so far is based on the questionable assumption that Q events were binomially distributed among subjects. In fact the scores in Table V were often attributable to a few subjects and the standard deviation of subject scores was often much higher than that predicted for a binomial distribution. For instance, one subject typed the word DEAR 13 out of 15 times at speeds satisfying the criterion.

Discussion

Given the assumptions of the latency model there is evidence that, for some subjects, certain words had privileges in transcription. If the test is acceptable then these privileges can be taken as evidence of response grouping. Anyone sceptical of the result can of course question the validity of the model itself. So far we have not attempted a mathematical fit to data, but a computer simulation indicates that the latency distributions at least can be quite well described by the model. In the simulation, latencies of the input system were assigned an exponential distribution and latencies of the output system a truncated normal distribution (disallowing negative latencies).

The tactic used here has been to assume that with trainee typists response grouping is at best a minority event. If this were not true then the test used would break down and it would be necessary to formulate a stronger, more explicit theory. It is planned to examine comparable data taken from professional typists to see whether response grouping is an occasional or essential feature of the high level skill.

We gratefully acknowledge financial support by the Science Research Council for this research, which is part of a study on serial skills.

References

BRYANT, W. L. and HARTER, N. (1899). Studies on the telegraphic language. *Psychological Review*, **6**, 345–75.

CRAIK, K. J. W. (1948). Man as an element in a control system. *British Journal of Psychology*, **38**, 142–8.

HERSHMAN, R. L. and HILLIX, W. A. (1965). Data processing in typing. *Human Factors*, **7**, 483–92.

LASHLEY, K. S. (1951). The problem of serial order in behaviour. In JEFFRESS L. A. (Ed.) *Cerebral Mechanisms in Behaviour*. New York: Wiley.

LEONARD, J. A. and NEWMAN, R. C. (1964). Formation of higher habits. *Nature*, **203**, 550–1.

SHAFFER, L. H. (1971). Attention in transcription skill. *Quarterly Journal of Experimental Psychology*, **23**, 107–13.

SHAFFER, L. H. and HARDWICK, J. (1968). Typing performance as a function of text. *Quarterly Journal of Experimental Psychology*, **20**, 360–9.

SHAFFER, L. H. and HARDWICK, J. (1970). The basis of transcription skill. *Journal of Experimental Psychology*, **84**, 424–40.

THOMAS, E. A. C. and JONES, R. G. (1970). A model for subjective grouping in typewriting. *Quarterly Journal of Experimental Psychology*, **22**, 353–67.

WICKELGREN, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behaviour. *Psychological Review*, **76**, 1–15.

Received 3 November 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 275–283

SHORT TERM FORGETTING IN THE ABSENCE OF PROACTIVE INTERFERENCE

A. D. BADDELEY AND DENISE SCOTT

Laboratory of Experimental Pyschology
University of Sussex

It has been claimed that the short-term forgetting shown by the Peterson technique is entirely due to proactive interference from prior experimental items. Two experiments investigated this by studying forgetting when prior items were avoided by testing subjects only once. Both experiments showed significant forgetting, although the degree of forgetting was less than with a multitrial procedure. On the basis of this and other results it is suggested that the Peterson technique comprises two components, a primary memory component which decays within 6 sec, and a more stable secondary memory component. Forgetting with the multitrial procedure is attributed principally to the need to use temporal retrieval cues to avoid confusion between successive items; longer retention intervals are associated with reduced temporal discriminability and hence poorer recall.

Brown (1958) and Peterson and Peterson (1959) demonstrated that sequences of items well within the immediate memory span will show marked forgetting within 15–20 sec if rehearsal is prevented by an interpolated task. Although this effect has been widely studied and is regarded as one of the central phenomena of short-term memory (STM), it has still no generally accepted explanation.

Both Brown (1958) and Peterson and Peterson (1959) suggested that the short-term forgetting observed reflects the fading of a memory trace. However, Keppel and Underwood (1962) showed that the characteristic rapid short-term forgetting curve does not occur for the first sequence of items a subject tries to remember. This first sequence shows little or no forgetting, suggesting that forgetting is largely due to proactive interference from prior experimental sequences. While there is no doubt about the importance of prior items in producing short-term forgetting, there is considerable disagreement about the underlying mechanism. Keppel and Underwood (1962) favour an interference theory interpretation which claims that the learning of an item (e.g. a CCC trigram) requires the unlearning of any previous items, these subsequently show spontaneous recovery and hence compete with the item the subject is trying to remember. The longer the retention interval, the greater the spontaneous recovery, the greater the competition and the lower the probability of recalling the correct item. However, this interpretation has difficulty in handling the demonstration by Peterson and Gentile (1965) that the amount of short-term forgetting decreases markedly when the interval between successive trials is lengthened. Conrad (1967) has also shown that the pattern of intrusion errors changes as a function of retention interval in a way which is difficult to reconcile with classical interference theory. He modifies trace decay theory to accommodate the PI effect by suggesting that prior items form a noisy background against which the subject must discriminate the fading trace of the item

to be recalled. On this interpretation, the first sequence presented showed littl or no forgetting in the Keppel and Underwood study because the absence of a nois background made the trace discrimination so easy that genuine trace decay wa masked by a ceiling effect. On subsequent trials, the "noise" provided by prio items exaggerates the effect of the fall in signal-to-noise ratio that constitutes th fading of the trace of the item to be recalled. If this is so it should be possible t show genuine forgetting of the first item presented, provided ceiling effects ar avoided. The following experiments aimed to discover whether such forgettin does occur.

Experiment I

Although a number of studies have failed to find any short-term forgetting o trial one (Keppel and Underwood, 1962; Cofer and Davidson, 1968; Turve Brick and Osborne, 1970) all these studies used trigrams. Performance on tri one was nearly perfect and hence it is possible that forgetting occurred but w masked by a ceiling effect. In a study using sequences of five or six word Houston (1965) showed very marked forgetting on the first trial, although th

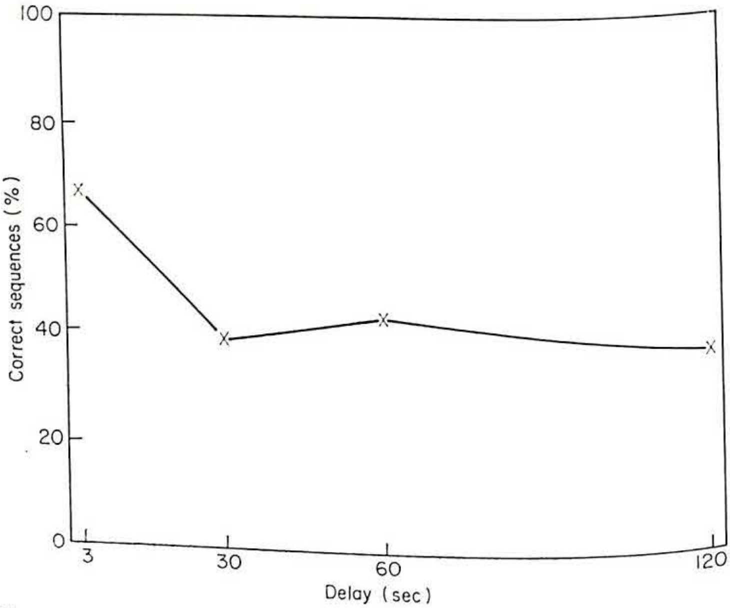

FIGURE 1. Short-term forgetting of a five-digit sequence when each subject is tested only o

occurred mainly in the first 1·5 sec, suggesting that it may have been due to so very short-term sensory store (Murray, 1968). Experiment I tried to avoid b ceiling effects and sensory memory effects by studying the retention of a si acoustically presented five-digit sequence over delays of 3, 30, 60 and 120 The sequence was presented auditorily at a rate of one digit per sec; during retention interval subjects copied letters also read out at a rate of one per sec, at the verbal signal "recall", attempted to write down the digit sequence in the o presented. A total of 152 young men were each tested once, in groups of 10

subjects, with an approximately equal number at each retention interval. The mean percentage of subjects reproducing the sequence completely correctly is shown in Figure 1. There is a clear drop between 3 and 30 sec ($\chi^2 = 4\cdot5$, $df = 1$ $P < 0\cdot05$), but no further forgetting occurred between 30 and 120 sec ($\chi^2 < 1$ in all cases.)

It seems then that forgetting does occur on trial one provided ceiling effects are avoided, indicating that PI from prior experimental items is not essential for short-term forgetting. It seems furthermore that such forgetting reaches asymptote within 30 sec, as is typical of multitrial studies (e.g. Peterson and Peterson, 1959). However, since there was no estimate of level of performance at points between 3 and 30 sec, it would clearly be premature to conclude that forgetting rates do not differ between single and multitrial studies. Experiment II attempts to investigate this by comparing the rate of forgetting for single and multitrial procedures.

Experiment II

Subjects tried to recall a single sequence of three, five or seven digits after a delay of 0–36 sec during which rehearsal was minimized by a letter copying task. Both digits and letters were read out at a rate of one per sec in time to a metronome. Subjects were given copying practice on one run each of five and ten letters; a digit sequence was then read out followed after a one sec gap by 0, 3, 6, 9, 18, or 36 letters to be copied, and the verbal signal "recall". They then attempted to write down the digit sequence in the order presented. Subjects were told in advance what length of digit sequence to expect but not what delay; recall was unpaced. Subjects were 424 undergraduates who were encouraged to volunteer by the offer of a small sum to a specified charity for each subject tested. They were tested in groups of three to eight in a caravan parked in the middle of the University of Sussex Campus, with a separate random number sequence being used for each group. Assignment of groups to a particular sequence length and delay was random, with the constraint that the number of subjects tested at each delay was roughly constant [for details of group size see Figure 2(a)].

For purposes of comparison a further group of subjects was tested using a multiple test design in which each subject was tested three times on each sequence length at each delay, making a total of 54 tests per subject. Two groups of 9 undergraduates were tested. Each group experienced a different sample of digit sequences and a different random order of presentation of sequences from the 18 conditions (three sequence lengths $\times$ six delays). Otherwise the procedure was the same as for the single test experiment.

Results

Figure 2(a) shows mean percentage of digits recalled in the correct serial position as a function of sequence length and delay for the single test conditions. Since the principal interest is in the slopes of the three lines, for the purpose of analysis, each point was expressed as a deviation from the mean for sequences of that length. Regression analysis indicated a significant overall tendency for performance to deteriorate as delay increased, $F(3,12) = 8\cdot9$, $P < 0\cdot01$, but no difference between

slopes for the three lines, $F(2,12) = 3{\cdot}53$, $P > 0{\cdot}05$. Analysis of the individual lines showed a significant slope for three digits, $F(1,4) = 18{\cdot}4$, $P < 0{\cdot}05$, and for five digits, $F(1,4) = 19{\cdot}4$, $P < 0{\cdot}05$, but not for seven digits, $F(1,4) = 1{\cdot}6$, $P > 0{\cdot}05$.

Performance on the multitrial experiment is shown in Figure 2(b). Analysis of Variance showed significant effects of list length, $F(2,80) = 155{\cdot}3$, $P < 0{\cdot}01$, delay, $F(5,80) = 50{\cdot}2$, $P < 0{\cdot}01$, and a significant interaction between list length and delay, $F(10,80) = 5{\cdot}8$, $P < 0{\cdot}01$. This interaction is almost certainly due to a ceiling effect which reduces the apparent rate of forgetting for sequences of length three, since there is clearly no difference in rate of forgetting between five- and seven-digit sequences.

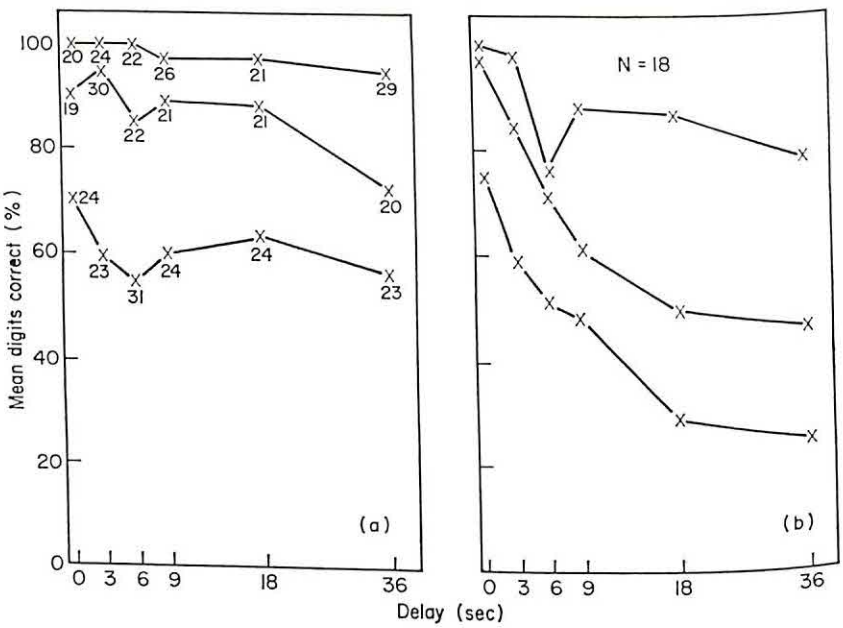

FIGURE 2. Short-term forgetting of sequences of 3, 5 and 7 digits (a) when each subject is tested only once, and (b) when subjects are tested repeatedly. Numbers on (a) refer to subjects contributing to the relevant data point.

Comparison with Figure 1(a) representing the performance of the subjects tested only once suggests that there is no difference in degree of learning as measured by performance at zero delay, but that multitrial subjects showed more forgetting, with performance declining to a lower level. These trends were evaluated statistically by comparing the two groups in terms of number of subjects totally correct on the immediate test and on the ultimate test after 36 sec. Comparison using the χ^2 test indicated no difference at zero delay, with significantly better performance by subjects tested only once at the 36 sec delay for sequences of three ($\chi^2 = 11{\cdot}5$, $df = 1$, $P < 0{\cdot}001$), and five digits ($\chi^2 = 8{\cdot}2$, $df = 1$, $P < 0{\cdot}01$). Comparison in terms of proportion of sequences correct rather than proportion of subjects making no errors gave the same result. With seven digits, level of performance in both conditions was too low to allow a meaningful comparison

on this measure. A comparison of forgetting rates in the two studies was made in terms of the percentage decrement between 0 and 36 sec shown for each sequence length. The performance of each subject in the multitrial condition was compared with the mean decrement shown in the single trial condition. The sign test showed a significantly greater degree of forgetting for sequences of length 5 and 7 ($P < 0{\cdot}001$ in each case) but not for length 3, in which 6 of the 18 multitrial subjects showed no forgetting whatsoever.

Discussion

It is clear from both experiments, that providing ceiling and floor effects are avoided, subjects will show reliable forgetting on a single trial. Forgetting with this technique can therefore not be attributed entirely to interference from prior experimental items. This does not, however, eliminate an interference theory explanation, since the theory can produce no less than three alternative explanations, extra-experimental interference, retroactive interference (Postman and Keppel, 1969, p. 348), and intra-sequence interference (Melton, 1963).

The extra-experimental interference hypothesis suggests that the learning of a sequence of items requires the unlearning of already existing language habits. These will spontaneously recover during the retention interval and will compete with the material to be recalled producing forgetting (Underwood and Postman, 1960). This hypothesis predicts faster forgetting of high frequency words since these have more and stronger competing prior associations. This hypothesis has not proved successful in LTM (Underwood and Ekstrand, 1966), although Turnage (1967) claims to have shown faster forgetting of high frequency words using the Peterson technique. However, his experiment is open to a number of criticisms, and subsequent experiments (Baddeley and Scott, 1971) have failed to produce any evidence in favour of the hypothesis either with a multitrial procedure or in an experiment in which each subject was tested only once.

An explanation in terms of retroactive interference from the interpolated task does not seem very probable either, since digits and letters are generally assumed to be sufficiently dissimilar to avoid interference effects (Wickens, Born and Allen, 1963). Furthermore, a recent study by Baddeley, Ecob and Scott (1970) suggests that RI as conceived by classical interference theory does not occur in STM. The effects that are traditionally cited as evidence for RI with the Peterson paradigm (e.g. Dale and Gregory, 1966) are shown to be strategy effects which occur when subjects are able to employ additional retrieval cues, but which disappear when such cues are not possible.

The intra-sequence interference explanation was devised by Melton (1963) to explain the apparent positive relationship between number of items to be remembered and rate of forgetting using the Peterson technique. The observed relationship is, however, attributable to ceiling effects with shorter sequences at short delays. When ceiling effects are avoided, no such interaction occurs (Baddeley, 1968; Baddeley and Scott, 1971). Furthermore, Experiment II of the present study clearly fails to show the predicted faster forgetting of longer sequences.

It appears therefore to be the case that, as Conrad's hypothesis would predict,

forgetting does occur in the absence of interference from prior items, and that such forgetting does not conform to the predictions of classical interference theory. Can the whole of forgetting with the Peterson technique therefore be attributed to trace decay? The fact that the multitrial condition led to more forgetting than the single trial is not a problem if one assumes, as Conrad suggests, that prior items exaggerate the effects of trace decay by forming a noisy background against which the subject must discriminate the trace of the relevant item. There is, however, one feature of single trial forgetting results that argues against this. In virtually all the single trial studies we have performed, there is a strong suggestion that forgetting approaches asymptote within approximately 5 sec. This effect is not clear-cut, since even with 20–30 subjects per point, forgetting curves are highly irregular. That it does occur with some consistency however, is indicated by Figure 3 which shows mean percentage of subjects recalling a sequence correctly in the six single trial conditions we have run, namely the three-, five- and seven-digit conditions from the present study (A, B and C), an unpublished replication of the five-digit condition in which digits were presented at a rate of two per sec (D), and the high and low frequency word sequence results (E and F) from the previously described study (Baddeley and Scott, 1971). There is a rapid fall in performance reaching asymptote within 5–6 sec in all except the three-digit condition for which ceiling effects prevent any conclusion about rate of forgetting in the first 6 sec. In all other conditions there is a sharp drop in performance over the first 5–6 sec after which there is a suggestion of improvement, followed by a subsequent tendency for performance level to decline gradually. This pattern of forgetting has also been found with the single test technique by Marcer (personal communication). If it can be established that the bulk of forgetting does indeed occur within the first 5 sec, then it would strongly suggest the existence of a factor other than trace decay in the results of multitrial Peterson technique experiments. Such results typically show a much later asymptote, as is illustrated in Figure 2(b). Analysis of forgetting between 6 and 36 sec using the sign test indicated significant forgetting for sequences of length 5 and 7 ($P < 0{\cdot}01$) though not for length 3. If performance on a single trial experiment reflects the strength of a fading trace, and if fading reaches asymptote within 6 sec, then further forgetting on Conrad's hypothesis can only occur if there is a gradual increase in background noise. Since such noise is provided by the traces of prior items, which will themselves have reached a stable level, it is not easy to see why there should be an increase in background noise.

Given the current, admittedly inadequate evidence, the most probable interpretation of forgetting with the Peterson technique would seem to be as follows: Presentation of an item lays down a trace in both a rapidly decaying primary memory (PM) store, and a more stable secondary memory (SM) store. The PM trace will decay within about 5 sec leaving the more stable SM trace. Such forgetting rates are consistent with estimates for the PM component in free recall (Glanzer, Gianutsos and Dubin, 1969) and in minimal paired-associate learning (Peterson, 1966). With a multitrial paradigm, further forgetting will occur due to confusion between successive items. Essentially, the subject has been given a series of items and is instructed to retrieve the most recent, that is to retrieve on the basis of a temporal cue (Yntema and Trask, 1963). The longer the delay, the

more difficult the temporal discrimination becomes and the greater the probability of error. An increase in the interval separating successive items improves recall (Peterson and Gentile, 1965; Loess and Waugh, 1967), since this increases the time difference between the item to be recalled and prior items, and thereby facilitates the temporal discrimination. A recent study by Turvey, Brick and

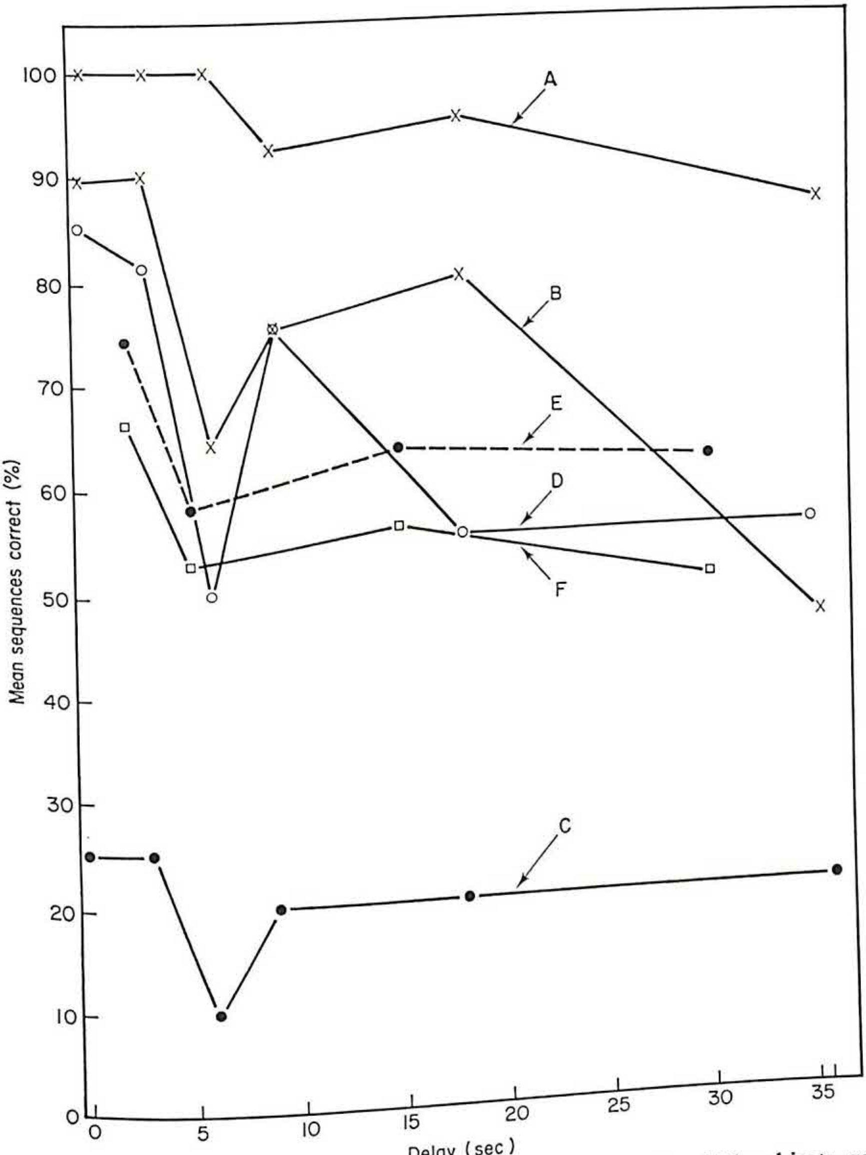

FIGURE 3. Short-term forgetting curves for a range of experiments in which subjects were tested only once. For details see text.

Osborne (1970) shows that prior item retention interval affects recall probability in a similar way. The longer the delay between presentation of the prior item and the current item, the easier the temporal discrimination and the greater the probability of correct recall.

We wish to suggest therefore that the Brown/Peterson technique comprises two separate components, a relatively small PM component which decays within the first 5 sec, and a larger SM component which shows relatively little direct forgetting. With a multitrial procedure, however, subjects are forced to rely on temporal cues to avoid confusion between successive items, an increase in retention interval will increase the difficulty of this discrimination and hence produce forgetting.

We are grateful to D. Baker for statistical advice and to the Medical Research Council for financial support. This paper was written while the first author was visiting the Center for Human Information Processing, University of California, San Diego, supported by N.I.M.H. Grant Number M.H. 15828.

References

BADDELEY, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology*, **20**, 240–64.

BADDELEY, A. D., ECOB, J. R. and SCOTT, D. (1970). RI in STM: A cognitive interpretation with implications for the role of acoustic and semantic coding. Paper presented at the 10th annual meeting of the Psychonomic Society, San Antonio, Texas.

BADDELEY, A. D., and SCOTT, D. (1971). Word frequency and the extra-experimental interference hypothesis in short-term memory. *Journal of Verbal Learning and Verbal Behaviour* **10**, 35–40.

BROWN, J. A. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, **10**, 12–21.

COFER, C. N. and DAVIDSON, E. H. (1968). Proactive interference in STM for consonant units of two sizes. *Journal of Verbal Learning and Verbal Behaviour*, **7**, 268–70.

CONRAD, R. (1967). Interference or decay over short retention intervals? *Journal of Verbal Learning and Verbal Behaviour*, **6**, 49–54.

DALE, H. C. A. and GREGORY, M. (1966). Evidence of semantic coding in short-term memory. *Psychonomic Science*, **5**, 75–6.

GLANZER, M., GIANUTSOS, R. and DUBIN, S. (1969). The removal of items from short-term storage. *Journal of Verbal Learning and Verbal Behaviour*, **8**, 435–47.

HOUSTON, J. P. (1965). Short-term retention of verbal units with equated degrees of learning. *Journal of Experimental Psychology*, **70**, 75–8.

KEPPEL, G. and UNDERWOOD, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behaviour*, **1**, 153–61.

LOESS, H. and WAUGH, N. C. (1967). Short-term memory and inter-trial interval. *Journal of Verbal Learning and Verbal Behaviour*, **6**, 455–60.

MELTON, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behaviour*, **2**, 1–21.

MURRAY, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, **78**, 679–84.

PETERSON, L. R. (1966). Short-term verbal memory and learning. *Psychological Review*, **73**, 193–207.

PETERSON, L. R. and GENTILE, A. (1965). Proactive interference as a function of time between tests. *Journal of Experimental Psychology*, **70**, 473–8.

PETERSON, L. R. and PETERSON, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, **58**, 193–8.

POSTMAN, L. and KEPPEL, G. (Eds.) (1969). *Verbal Learning and Memory*. Harmondsworth: Penguin Books.

TURNAGE, T. W. (1967). Unit sequence interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, **6**, 61–5.

TURVEY, M. T., BRICK, P. and OSBORNE, J. (1970). Proactive interference in short-term memory as a function of prior item retention interval. *Quarterly Journal of Experimental Psychology*, **22**, 142–7.

UNDERWOOD, B. J. and EKSTRAND, B. R. (1966). An analysis of some shortcomings of the interference theory of forgetting. *Psychological Review*, **73**, 540–9.

UNDERWOOD, B. J. and POSTMAN, L. (1960). Extra-experimental sources of interference in forgetting. *Psychological Review*, **67**, 73–95.

WICKENS, D. D., BORN, D. G. and ALLEN, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, **2**, 440–5.

YNTEMA, D. and TRASK, F. (1963). Free recall as a search process. *Journal of Verbal Learning and Verbal Behaviour*, **2**, 65–74.

Received 13 *November* 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 284–295

SHADOWING AND MONITORING FOR SELECTIVE ATTENTION

GEOFFREY UNDERWOOD AND NEVILLE MORAY†
Department of Psychology, University of Sheffield

Although shadowing has been assumed to hold attention strongly (e.g. Hochberg, 1970) this has not been demonstrated, and the incidence of errors and intrusions suggest that it is not as powerful a technique as might have been supposed. An alternative, simpler, technique would be monitoring, but Kahneman (1970) maintains that selectivity fails in the absence of a continuous response. The experiments here compare the shadowing and monitoring tasks when attentional selectivity was required, and when it was not required, and indicate that similar attentional strategies operate during the monitoring and shadowing of brief messages. The shadowing paradigm is criticized in terms of an interaction between the relative resemblance of the stimuli and the shadowing voice. It is not demonstrated that an absolute interaction takes place, but more importantly that relatively greater interference is apparent when the shadower's voice and the stimuli to be shadowed are similar than when they are distinct. It is suggested that low target detections in the unattended message may be an artifact of the processing requirements of the shadowing task.

Introduction

Since shadowing was first used as an experimental technique by Cherry in 1953 it has repeatedly been used in investigations of the parameters of selective attention as it is "one method that will both force the subject to attend to the primary message and test his reception of it" (Hochberg, 1970). Most of the data produced have been collected from experiments motivated by competition between the theories of selective attention, and although these experiments generally rely upon the shadowing procedure it has not been shown to hold attention consistently. Moray (1969) has argued that any modification of the direction of attention is reflected in a decrease of shadowing accuracy, but this can only be true if correct shadowing guarantees maximum attention to the shadowed message.

In the dichotic situation shadowing errors and intrusions from the secondary message appear to be principally a function of the type of material presented. Reports of intrusions, however, may indicate that at least in some circumstances shadowing does not hold the listener's attention onto a particular message, but merely indicates the direction of attention. Even if intrusions are due to the difficulty of spatially discriminating between the two signals (Treisman and Riley, 1969), the incidence of intrusions serves to show that the shadowing response is a function of input selection. Shadowing is only possible if material is processed following spatial analysis; it may motivate this differentiation but it cannot control it.

† Present address: Department of Psychology, Scarborough College, University of Toronto, Canada

A number of recent experiments (Moray and O'Brien, 1967; Shaffer and Hardwick, 1969) have abandoned the shadowing task, and subjects have been instructed simply to listen to (monitor) a relatively long message. That instructions to monitor a single source in a dichotic presentation are sufficient for this to occur is not established, and Kahneman (1970) concludes that in the selective monitoring situation "selective attention is far from optimal". Presumably the arguments about discriminating between signals during shadowing hold for monitoring, and when Kahneman's monitoring subjects report items from the rejected channel despite a penalty, they may well be reflecting a difficulty to discriminate between signals. If we assume that shadowing efficiency is correlated with the direction of attention, then shadowing can be used as a measure of which message is being primarily attended, but during monitoring we cannot be sure whether the subject is attending to the primary and/or secondary message as no overt continuous response is made. Monitoring is essentially a passive task, and unless the signals are presented at a very fast rate the listener requires little effort to analyse the components fully. During shadowing, however, the listener is making a continuous verbal response to a continuously changing verbal stimulus and has his information processing capacity taxed to a much greater extent than in the monitoring situation. If the listener has a fixed amount of capacity available (Moray, 1967), and different amounts of the total capacity are taken by monitoring and by shadowing, then different residual amounts are available for the experimental target detection task. The degree to which the secondary message can be processed is thus not only a function of the capacity required for primary message analysis and the monitoring of the attention sharing process (Taylor, Lindsay and Forbes, 1967), but also of the amount of capacity required by the experimental technique. Peterson (1969) investigated performance on an anagram solving task in the absence and presence of various other verbal activities (including shadowing) and similarly concluded that attention allocated to the primary task is restricted in that attention must be shared with the serial activity. The shadower's own voice, absent during monitoring, may act as a third signal source, causing interference with the perception of the incoming messages. If, as Norman (1969) and Glucksberg and Cowen (1970) indicate, the unattended message reaches a buffer storage, then one way in which shadowing could produce false detection rates is by the shadowing voice interfering with transient echoic traces which are awaiting processing, in a way similar to Crowder and Morton's (1969) stimulus suffix. Kroll *et al*. (1970) have used the disturbing and capacity consuming nature of the shadowing task to advantage, in a short-term memory study.

Direct comparison between shadowing and monitoring experiments is not possible because relevant variables (e.g. type of material, instructions to subjects) have not been held constant. The aim here is to investigate the relationships between shadowing, monitoring and attentional selectivity.

Experiment I

As the first step in checking the effectiveness of instructions to monitor one message of a dichotic array detection performance in the absence of the factor of

voluntary attentional control was collated over monaural and dichotic trials. In this experiment all of the targets to be detected were in the primary message.

Method

Stimulus material

16 lists each of 16 randomly selected letters were recorded in a male voice (MV) at the rate of 2·5 letters/sec for monaural presentation from Channel 1 of a stereo tape-recorder. Similarly, 16 dichotic lists were prepared, and inspected for synchrony and consistency using an oscilloscope. Onset synchrony of the pairs of items was judged to be within 50 msec. The imminence of the lists was indicated by the words "ready, go" recorded on Channel 1. One randomly selected digit, to serve as a detection target was inserted in each Channel 1 list. The digit did not occur during the first or last four items. In half of each category of recordings the digit was spoken in the same voice (SV) as the letters, and in half in a distinctly different voice (DV), a female voice (FV). All lists were presented through stereo earphones. The only difference between the monaural and dichotic situations, therefore, is that in the latter case a synchronized redundant list of letters is present in the unattended ear. Several monaural and dichotic sets of letters without targets were recorded for the purpose of shadowing practice.

Subjects

Subjects in all experiments described in this report were drawn on a selection without replacement basis from the same volunteer subject pool, which was composed of undergraduate and research students of the University. In this experiment 20 males and 20 females were used.

Procedure

Ten males and ten females were first given shadowing practice with monaural lists until they were reasonably proficient. Half of each group of subjects listened with their right ears and half with their left ears. The 16 monaural lists were divided into four groups such that subjects monitored 8 lists and shadowed 8 lists, 4 lists in each case containing an SV digit and 4 a DV digit. In each trial subjects were told of the voice type of the target digit, and that they were to repeat the digit as soon as they heard it. In the shadowing trials subjects were told that they must shadow all of the list until they heard the target, at which point they should stop shadowing and say the digit.

Subjects' continuous responses were counterbalanced over an ABBA design. Half of the right ear subjects and half of the left ear subjects monitored the first block of four trials, shadowed the following eight trials, and monitored the last four trials. The other subjects shadowed the first block, monitored two blocks, and shadowed the last block. Each block consisted of two SV trials and two DV trials.

After presentation of the monaural lists shadowing practice with dichotic stimuli was given, and at this time subjects individually subjectively equated the loudness of the two signals. The 16 experimental dichotic trials were presented, grouped and counter balanced as in the monaural situation.

As would be expected, the monitoring task proved to be extremely simple for subjects and so to make it slightly more difficult the remaining 10 males and 10 females were used in a repeat of the procedure in the presence of masking "white" noise. White noise, from independent sources, at a level of 65 dB was introduced to both earphones in all conditions. The signal-to-noise ratio was approximately −3dB. Subjects subjectively equated the volume of the rejected ear noise to that of the fixed volume accepted ear noise.

Results

In each trial the subject had to respond with the correct digit to score a successful detection. No incorrect digits were given in any SV target trial, and incorrect

digits or reports of subjects noticing a change of physical characteristic (PCO) in the DV trials were recorded separately. All subjects monitoring the stimuli without white noise detected targets with ease, and the results from this condition were discarded. The percentages of targets detected in the remaining conditions are given in Table I.

The results from the shadowing trials of the noise and no noise conditions, and those from the shadowing and monitoring trials of the noise condition were submitted to two separate analyses of variance.

TABLE I

	Same voice (MV)				Different voice (FV)			
	Monaural		Dichotic		Monaural		Dichotic	
	M	S.D	M	S.D	M	S.D	M	S.D
No noise								
Shadowing								
Males:	77·5	14·2	62·5	17·7	90·0 (90·0)	17·5 (17·5)	95·0 (97·5)	10·5 (7·9)
Females:	82·5	16·8	77·5	18·4	80·0 (85·0)	10·5 (12·9)	90·0 (95·0)	17·5 (15·8)
Noise								
Shadowing								
Males:	30·0	22·9	30·0	22·9	42·5 (75·0)	20·5 (20·4)	80·0 (92·5)	18·4 (12·0)
Females:	50·0	20·0	47·5	21·8	55·0 (70·0)	21·8 (16·6)	67·5 (75·0)	16·8 (19·7)
Monitoring								
Males:	50·0	20·0	42·5	16·8	80·0 (85·0)	7·9 (12·9)	77·5 (95·0)	21·8 (10·5)
Females:	47·5	21·8	37·5	22·9	70·0 (77·5)	19·7 (21·8)	80·0 (95·0)	18·4 (10·5)

Percentages of targets correctly detected in Experiment I. Background items in a MV, targets in the same (MV) or a different (FV) voice. Figures in brackets refer to combined percentages of correct semantic detections and reports of physical change only (PCO's).

During shadowing, the incidence of white noise had a general effect on the detection rates ($F = 44{\cdot}6$, $df = 1{,}36$, $P < 0{\cdot}001$). The voice type of target was also influential, SV digits detected on 58·8% and DV digits on 83·3% of all occasions ($F = 42{\cdot}2$, $df = 1{,}36$, $P < 0{\cdot}001$). Noise was observed to disturb SV more than DV target detections ($F = 5{\cdot}2$, $df = 1{,}36$, $P < 0{\cdot}05$). An interesting interaction emerged between the voice type of target over monaural and dichotic trials ($F = 6{\cdot}9$, $df = 1{,}36$, $P < 0{\cdot}025$). DV targets had a detection advantage in monaural trials (DV: 80%; SV: 62·5%), but this advantage was increased in dichotic trials (85·6%; 55%). The increase in DV target detections may be facilitated by additional physical information at the instant of presentation producing a mismatch of frequencies (Durlach, 1963), and therefore easier detection,

whereas the decrease in SV detections may be due to the the additional masking provided by noises of similar frequency on the rejected channel.

When the data were re-analysed, including PCO reports as successful target detections, all of these findings were confirmed.

Comparison of all conditions of the white noise trials established that monitoring the stimuli (66%) provided more detections ($F = 5{\cdot}6$, $df = 1{,}18$, $P < 0{\cdot}05$) than did shadowing (50%). The voice type of the target remained a highly significant variable, and the interaction between this factor and that of the monaural/dichotic presentation was also confirmed.

Considering PCO reports also as DV detections eliminated the monitoring/shadowing difference, indicating that the additional masking of the subject's own voice during shadowing does not totally mask the DV target, but does mask it sufficiently well to make semantic content extraction more difficult than when monitoring.

The interference of the shadowing voice is also demonstrated by a three way interaction involving response activity, voice type of target, and sex of subjects ($F = 5{\cdot}2$, $df = 1{,}18$, $P < 0{\cdot}05$). All subjects monitoring produce similar differences between responses to SV (44·4%) and DV (88·1%) digits. Males shadowing also produce this difference (30%: 81·8%), but females shadowing do not derive such a great advantage from DV targets (48·3%: 72·5%). The female shadowing voice is apparently interfering less than the male shadowing voice with SV (i.e. MV) targets, whereas male subjects have the advantage with DV (i.e. FV) targets. The shadowing data of the no noise condition, when analysed separately, also illustrate this interaction ($F = 4{\cdot}9$, $df = 1{,}18$, $P < 0{\cdot}05$). Thus, when shadowing, the type of the subjects own voice selectively interferes with the type of target voice to decrease detection probability when the two voices are similar.

Experiment Ia

Although it might be argued that in the previous experiment selective interference, during shadowing, was shown with MV and FV targets by male and female subjects' voices, this has only been demonstrated for MV background items with MV and FV test items. To establish the generality of the interaction this experiment used FV background items with FV and MV test items.

Method

Stimulus material and procedure

Sixteen dichotic lists of sixteen letters each with a digit substituted for a Channel 1 letter were recorded by a female speaker. The lists were comparable to those used in the dichotic trials of Experiment I, and were checked for synchrony and consistency as before. The procedure was essentially similar to that described above. White noise was not used, and subjects only shadowed the lists. Each subject was thus presented with 8 DV and 8 SV targets, all of which had channel incidence certainty.

Subjects

Ten males and ten females, drawn from the same subject pool as those used earlier, were used in this experiment.

Results

Detection percentages are given in Table II. Response differences to SV targets (72·5%) and DV targets (91·25%) were again observed ($F = 54.7$, $df = 1{,}18$, $P < 0.001$), as was the interaction between sex of shadower and voice type of

TABLE II

	Same voice (FV)		Different voice (MV)	
	M	S.D.	M	S.D.
Males:	75·0	8·3	87·5 (93·75)	5·9 (8·8)
Females:	70·0	12·1	95·0 (98·75)	6·4 (7·9)

Percentages of targets correctly detected in Experiment Ia. Background items in a FV, targets in FV or MV. Detections, including reports of PCO's are given in brackets.

target ($F = 4.8$, $df = 1{,}18$, $P < 0.05$). This indicates that subject's own voice does consistently interfere with the shadowed stimuli, and impairs perception of messages of the same voice type more than those of the different voice type.

Experiment II

The comparison of the monitoring and shadowing modes of response was continued here, with the addition of the factor of uncertainty of target incidence. The design thus approached the traditional selective auditory attention situation.

Method

Stimulus material

Thirty-two dichotic lists of random letters, comparable to those used in the previous experiments, were recorded by a male speaker with the target digit occurring pseudo-randomly in either channel such that half were in each channel. Of the 16 accepted channel lists, half of the inserted targets were recorded in the SV as the letters and half in a FV (DV). Similarly with the rejected channel lists. Thus there were 8 lists in each stimulus category.

Subjects

Twelve male and twelve female subjects were used, drawn from the source of subjects used previously.

Procedure

Half of each group of subjects performed the experiment with white noise added to the stimuli, as in Experiment I. Channel volumes and, where appropriate, noise volumes were subjectively equated before practice was given at shadowing. Half of each of the two groups of males and females were presented with the Channel 1 (accepted) message to their right ears, and half to their left. Subjects were told that the target digit may occur in either ear, but that they were not to miss any targets occurring in their accepted ear. Of the 8 dichotic lists in each stimulus condition, half were monitored and half shadowed. A counter-balanced response activity design, described in Experiment I, was employed.

Results

The percentages of targets detected in each of the conditions are presented in Table III. Analysis of variance on the detection performances again revealed that noise had a significant effect ($F = 45{\cdot}1$, $df = 1{,}22$, $P < 0{\cdot}001$), although no interactions involved this factor. Other highly significant main factors were that monitoring resulted in more detections than shadowing ($F = 45{\cdot}9$, $df = 1{,}22$, $P < 0{\cdot}001$); DV targets were more easily detected than SV targets ($F = 148{\cdot}9$, $df = 1{,}22$, $P < 0{\cdot}001$); and accepted channel targets were more often detected than rejected channel targets ($F = 108{\cdot}6$, $df = 1{,}22$, $P < 0{\cdot}001$). Shadowing was found to reduce the detection of SV targets (as compared with the detection of DV targets) more than did monitoring ($F = 11{\cdot}8$, $df = 1{,}22$, $P < 0{\cdot}01$), again illustrating the effect of the shadowing voice on the detection task. SV targets were also detected proportionately less often than DV targets when in the rejected rather than the accepted channel ($F = 39{\cdot}3$, $df = 1{,}22$, $P < 0{\cdot}001$).

TABLE III

	Same voice (MV)				Different voice (FV)			
	Accepted channel		Rejected channel		Accepted channel		Rejected channel	
	M	S.D.	M	S.D.	M	S.D.	M	S.D.
No noise								
Shadowing								
Males:	66·7	14·5	4·2	10·2	95·8 (100)	10·2 (0·0)	83·3 (95·8)	12·9 (10·2)
Females:	62·5	13·6	16·7	12·9	83·4 (87·5)	12·9 (13·6)	66·7 (75·0)	14·5 (15·8)
Monitoring								
Males:	83·3	12·9	41·7	20·4	100 (100)	0·0 (0·0)	83·3 (91·7)	12·9 (12·9)
Females:	95·8	10·2	41·7	20·4	95·8 (100)	10·2 (0·0)	91·7 (95·8)	12·9 (10·2)
Noise								
Shadowing								
Males:	37·5	13·6	0·0	0·0	70·8 (83·4)	10·2 (12·9)	45·8 (62·5)	20·4 (20·9)
Females:	33·4	20·4	4·2	10·2	75·0 (75·0)	15·8 (15·8)	66·7 (70·8)	14·5 (10·2)
Monitoring								
Males:	54·2	20·9	12·5	13·6	85·4 (95·8)	12·9 (10·2)	62·5 (95·8)	20·9 (13·6)
Females:	87·5	13·6	16·7	12·9	79·2 (91·7)	18·8 (12·9)	62·5 (83·3)	20·9 (12·9)

Percentages of targets correctly detected in Experiment II. Background items in a MV, targets in MV or FV. Detections, including reports of PCO's given in brackets.

Only when PCO reports were included as successful responses did the interaction between voice type of target and type of shadowing voice become evident $F = 6{\cdot}1$, $df = 1{,}22$, $P < 0{\cdot}025$); this interaction held whether subjects shadowed

or monitored. The three way interaction involving response activity, sex of subjects, and target voice failed to reach significance ($F = 3{\cdot}2$, $df = 1{,}22$).

Experiment IIa

To investigate further the interaction between the shadower's voice and the voice of the stimulus an experiment comparable to Experiment Ia was conducted.

Method

Stimulus material and procedure

Thirty-two dichotic lists were prepared, to duplicate the conditions of Experiment II, with the exception that the background letters were spoken by the same female voice as in Experiment Ia. Targets were in either a FV or MV. White noise was not used, and the instructions were as in Experiment II.

Subjects

Ten males and ten females were used, drawn from the source of subjects used earlier.

Results

Target detection percentages are given in Table IV. Analysis of variance

TABLE IV

	Same voice (FV)				Different voice (MV)			
	Accepted channel		Rejected channel		Accepted channel		Rejected channel	
	M	S.D.	M	S.D.	M	S.D.	M	S.D.
Shadowing								
Males:	77·5	18·4	12·5	13·2	72·5 (80·0)	18·4 (15·8)	57·5 (65·0)	16·8 (17·5)
Females:	70·0	19·7	7·5	12·1	87·5 (90·0)	17·6 (12·9)	77·5 (82·5)	18·4 (16·8)
Monitoring								
Males:	92·5	12·1	55·0	20·4	92·5 (97·5)	12·1 (7·6)	90·0 (95·0)	12·9 (10·5)
Females:	90·0	12·9	55·0	15·8	97·5 (100)	7·9 (0·0)	90·0 (95·0)	12·9 (10·5)

Percentages of targets correctly detected in Experiment IIa. Background items in a FV, targets in FV or MV. Detections, including reports of PCO's given in brackets.

again indicated highly significant differences between shadowing and monitoring, between SV and DV targets, and between accepted and rejected channel targets. The voice type of target again led to differential detection during monitoring and shadowing, and in their accepted and rejected channel positioning.

Targets were more often detected when the subject's own voice was dissimilar to the target voice than when the subject's voice was similar. That is, male subjects detected fewer MV targets, and more FV targets, than did female subjects. Although this interaction held whether subjects were shadowing or monitoring ($F = 6{\cdot}1$, $df = 1{,}18$, $P < 0{\cdot}025$), a further interaction suggests that the presence

of a subject's voice, during shadowing, gives rise to differential detection rates over target voice types more than does the passive monitoring response mode ($F = 4{\cdot}6$, $df = 1{,}18$, $P < 0{\cdot}05$).

For the results produced by this experiment, all of these differences were demonstrated when considering reports of PCO's also as successful detections of DV targets.

Discussion

The difference between SV and DV target detections, demonstrated in all of these experiments, confirm the general trends indicated by Treisman and Riley (1969), although in no case was a 99% detection rate of DV targets in the rejected channel reported here. According to their comments after the experiment subject's did tend to rely upon the detection of the physical change before extracting the verbal content, and the success rates therefore only indicate their ability to detect that change and consequently review the memory trace to find which digit the signal corresponded to. As far as the results demonstrate a difference in SV and DV detections the significance is limited, for Cherry's initial experiments reported that subjects usually notice a change of voice in the rejected ear. The present results are perhaps of more interest to investigations of memory processes, in that it is demonstrated that verbal content cannot always be determined after a DV is detected.

There were a number of interactions between the pitch of the shadower's voice and the pitch of the recorded target voice, and although these trends were not consistent enough for predictions to be derived, they do indicate that variability of target detection can in part be accounted for by the relationship between the target voice and the subject's voice. The shadowing voice appears to hinder subjects differentially—not even all males or all females will be affected consistently, for the range of pitch of voices within each sex varies as much as the range between the sexes. The addition of noise from the subject's own mouth tends to mask the incoming stimulus making analysis difficult but not impossible, otherwise the shadowing response would break down. It is possible that the occurrence of this sex difference and target voice interaction during monitoring can be explained in terms of ideo-motor mechanisms (Greenwald, 1970). If speech is perceived with the aid of the reconstruction of auditory images, then images of stimulus items in hese experiments, which would presumably be related to the listener's perceptions of their own voices, may be causing subsequent interference with stimulus items in a voice similar to that of the images.

The comparison of monitoring and shadowing trials in Experiments I and Ia suggested that monitoring enabled more efficient detection of all targets than did shadowing. Assuming that spatial localization of the primary and secondary messages was possible, then there was no attentional factor present in these experiments: subjects were certain about which ear to attend to in order to optimize detection. The differences between detections are thus due to differential response procedures during monitoring and shadowing. Output initiation and output monitoring require processing capacity during shadowing which can be delegated to

the target detection task in the monitoring mode. Peterson (1969) has similarly demonstrated the adverse effect of concurrent shadowing on an information processing task. The difference in the amounts of capacity available in the shadowing and non-shadowing conditions, and masking from shadowing voices leads to a performance decrement on the primary task in the more active response situation. Although subjects may have been dividing (Moray, 1960; Treisman, 1969) or multiplexing (Lindsay, 1970) rather than focusing their attention in the monitoring situations this strategy would not provide any detection advantages; indeed, it would be disadvantageous in that the primary signal in the dichotic situations would be masked by the secondary signal more when multiplexing than when focusing. Hockey (1970*a*, *b*) has demonstrated that loud noise increases attentional selectivity, and if this result is applicable to the dichotic listening situation it may be that this shift in behaviour is from a multiplexing strategy (quiet) to one of focusing (noise). If this is the case then a number of predictions can be made concerning the present data. If selectivity is altered by the incidence of noise in the monitoring conditions of Experiment I, then any masking influence of an irrelevant stimulus (the redundant list in the dichotic trials) should have more effect in the quiet condition when the strategy is multiplexing. Any influence of the masking stimuli should appear by comparison with the monaural trials. The influence of noise should be to facilitate focusing, to the exclusion of the irrelevant masking stimuli. In the multiplexing paradigm the monaural trials should provide more detections than the dichotic trials, and with the addition of noise and the shift to a focusing strategy there should be little difference between monaural and dichotic trials. Comparison of monitoring over monaural and dichotic trials without noise is not pertinent, as both presentations enabled simple detection of all targets. With the addition of noise monaural detections were reduced to 61·9%, and dichotic detections to 59·3%. The apparent absence of any interaction here suggests that noise is not differentially influencing attentional strategy over monaural and dichotic trials, which in the noise condition is taken to be one of focusing.

Experiments II and IIa introduced the factor of target incidence uncertainty, and the use of Hockey's findings can again indicate attentional strategy. If multiplexing is employed during monitoring, and focusing during shadowing, in the quiet conditions, and if focusing during monitoring and shadowing is employed in the noisy conditions, then an interaction should be observed between the response mode and the noise factor. This interaction was not apparent ($F = 0{\cdot}38$, $df = 1{,}22$), suggesting either that attentional strategy makes little difference to detection performances, or that subjects' strategies were similar during monitoring and shadowing. If subjects were multiplexing whilst monitoring without noise, then enhanced selectivity induced by the addition of noise should lead to greater impairment of rejected message detections than of accepted message detections, whereas the proportions of accepted and rejected message targets detected during shadowing should remain constant. The absence of a noise × response mode × target channel interaction ($F = 2{\cdot}77$, $df = 1{,}22$) may suggest the absence of a change in attentional strategy.

The experiments support the view that subjects' attention can be focused on a

given message in a dichotic array by instructions to monitor that signal. An explanation of the similarity of the patterns of response observed here is that subjects were voluntarily directing their attention as instructed, during monitoring and shadowing. Consequently, for short lists (as opposed to the longer stimulus presentations of Moray and O'Brien, 1967; Shaffer and Hardwick, 1969; and Kahneman, 1970) Kahneman's hypothesis that selectivity fails during monitoring would be questioned. An alternative explanation is that all subjects were multiplexing while monitoring and shadowing in all conditions. This would imply that Hockey's principle cannot be applied to the dichotic listening situation, and that shadowing does not facilitate focusing of attention.

The reliability of the technique of shadowing is discredited by the appearance of voice interactions between stimulus and shadower in these experiments. The low rejected message detections traditionally obtained may not be so much a reflection of diverted attention but of the high information processing load of the shadowing task. Consistently higher detections were obtained in the monitoring trials of these experiments, and if the attentional strategies were similar in both response modes, as was indicated, then it may be the case that the use of the monitoring mode will yield more accurate estimates of unattended message analysis than has the use of shadowing.

This work was supported by an S.R.C. Research Studentship to G.U. We are grateful to Miss S. A. Oakley for help with recording the material, to Mr. J. K. Clarkson for helpful discussions, and to Dr J. M. Wilding for comments on an earlier draft of the manuscript.

References

CHERRY, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America*, **23**, 975–9.

CROWDER, R. G. and MORTON, J. (1969). Precatagorical acoustic storage (PAS). *Perception and Psychophysics*, **5**, 365–73.

DURLACH, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *Journal of the Acoustical Society of America*, **35**, 1206–18.

GLUCKSBERG, S. and COWEN, G. N. (1970). Memory for non-attended auditory material. Paper read at the London Conference of the Experimental Psychology Society.

GREENWALD, A. G. (1970). Sensory feedback mechanisms in performance control: with special reference to the ideo-motor mechanism. *Psychological Review*, **77**, 73–99.

HOCHBERG, J. (1970). Attention, organization, and consciousness. In Mostofsky, D. I. (Ed.), *Attention: Contemporary Theory and Analysis*. New York: Appleton-Century-Crofts.

HOCKEY, G. R. J. (1970*a*). Effect of loud noise on attentional selectivity. *Quarterly Journal of Experimental Psychology*, **22**, 28–36.

HOCKEY, G. R. J. (1970*b*). Signal probability and spatial location as possible bases for increased selectivity in noise. *Quarterly Journal of Experimental Psychology*, **22**, 37–42.

KAHNEMAN, D. (1970). Remarks on attention control. In SANDERS, A. F. (Ed.), *Attention and Performance III*. Amsterdam: North Holland.

KROLL, N. E. A., PARKS, T., PARKINSON, S. R., BIEBER, S. L. and JOHNSON, A. L. (1970). Short-term memory while shadowing: recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, **85**, 220–4.

LINDSAY, P. H. (1970). Multichannel processing in perception. In Mostofsky, D. I. (Ed.), *Attention: Contemporary Theory and Analysis*. New York: Appleton-Century-Crofts.

MORAY, N. (1960). Broadbent's filter theory: postulate H and the problem of switching time. *Quarterly Journal of Experimental Psychology*, **12**, 214–20.
MORAY, N. (1967). Where is capacity limited? A survey and a model. In Sanders, A. F. (Ed.), *Attention and Performance*. Amsterdam: North Holland.
MORAY, N. (1969). *Attention: Selective Processes in Vision and Hearing*. London: Hutchinson.
MORAY, N. and O'BRIEN, T. (1967). Signal-detection theory applied to selective listening. *Journal of the Acoustical Society of America*, **42**, 765–72.
NORMAN, D. (1969). Memory while shadowing. *Quarterly Journal of Experimental Psychology*, **21**, 85–93.
PETERSON, L. R. (1969). Concurrent verbal activity. *Psychological Review*, **76**, 376–86.
SHAFFER, L. H. and HARDWICK, J. (1969). Monitoring simultaneous auditory messages. *Perception and Psychophysics*, **6**, 401–4.
TAYLOR, M. M., LINDSAY, P. H., and FORBES, S. M. (1967). Quantification of shared capacity processing in auditory and visual discrimination. In SANDERS, A. F. (Ed.), *Attention and Performance*. Amsterdam: North Holland.
TREISMAN, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, **76**, 282–9.
TREISMAN, A. M. and RILEY, J. G. A. (1969). Is selective attention selective reception or selective response? A further test. *Journal of Experimental Psychology*, **79**, 27–34.

Received 2 December 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 296–303

EFFECTS OF CHANGES IN THE INTENSITY OF WHITE NOISE ON SIMULTANEITY JUDGEMENTS AND SIMPLE REACTION TIME†

A. J. SANFORD

M.R.C. Applied Psychology Unit, Cambridge‡

On the basis of earlier work and informal observation it was suspected that the effect of loudness on simple reaction time (RT) could not be accounted for by changes in the time it takes the subject to hear the stimulus. Two experiments are described in which an increment in the level of background random noise is presented to the subject. The effect of increment size on RT and on a simultaneity judgement are investigated using a range of increments from just above difference threshold to moderately loud and clear. The difference in the size of loudness effects in the two tasks lends some support to a model which explains the influence of loudness on RT largely in terms of latency of response initiation.

Introduction

This paper is an investigation of the well-known relationship between simple reaction time and stimulus intensity. It has long been established that reaction time becomes shorter as stimulus strength increases (e.g. Chocholle, 1945) and concern here is with the possible explanation of the effect.

On a very simple line of reasoning it may be supposed that the subject reacts by pressing as soon as he has heard the stimulus and that the lag due to reduced loudness is due to a lag in the time it takes for the stimulus to be detected. Roufs (1963) has argued this point of view for the visual modality by comparing the size of the effect of intensity on reaction time with the size of the effect in the point of subjective simultaneity in psychophysical experiments, and finding that there was no difference between them. Thus Roufs has shown that the judgement of the simultaneity of two stimuli (which depends on the time of detection of the two stimuli) is lawfully related to the strengths of the stimuli, such that the dimmer the stimulus, the longer it takes the subject to detect it. In addition, the agreement between the size of the intensity effects measured under RT and simultaneity conditions indicates that RT-intensity effects are mediated through changes in the time it takes the subject to detect stimulus.

Rouf's findings are support for only one possible view of the RT process. An alternative view would be that neural activity resulting from stimulus presence

†Based on data from a thesis submitted to the University of Cambridge in partial fulfilment of the requirements for the degree of Doctor of Philosophy. I am grateful to Dr D. E. Broadbent for his supervision and to the Royal Navy for the provision of subjects.

‡Present address: Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland.

simply increases a tendency to respond which develops during the warning interval, the two (preparatory tendency and stimulus activity) adding and passing a response threshold so that a response is initiated. Such an argument does not necessarily involve any statement concerning the relationship between lag in perception and reaction time.

Interest in the problem here began with the observation that subjects know they react more slowly to quiet signals than loud ones, a result formally demonstrated in an experiment where subjects had to rate their own reaction speeds on a trial-by-trial basis (Sanford, 1970). Taken at face value this is curious—it implies that subjects are *aware* of a lag between detecting and responding to a quiet signal, suggesting that in the *auditory* case detection time is not affected by intensity as much as is reaction time. [Roufs (1963) has reached a similar conclusion without formal report.]

The experiments described here are an attempt to clarify the issue of the effects on intensity on reaction time and simultaneity. On the basis of results of Sanford (1970) it is anticipated that if there is any effect of auditory intensity on judgements of simultaneity such an effect will be smaller than the effect of intensity on reaction time.

There is now a considerable body of evidence that the size of the loudness effect in reaction time is not fixed but varies for the same intensity difference depending on the exact conditions of testing. Thus temporal uncertainty (Kellas, Baumeister and Wilcox, 1969) sequence effects (John, 1967*a*), proportion of loud to soft stimuli (Sanford, 1969) and warning stimulus intensity (Kohfield, 1969) are just some of the variables which influence the size of the intensity effect.

It is usual to summarize such effects by assuming that loud signals provide a criterion amount of "evidence" faster than quiet ones and that this gives rise to the intensity effect. The larger this criterion amount becomes the larger will be the difference between RTs to loud and soft stimuli since the evidence accumulation functions diverge (Grice, 1968; John, 1967*b*).

On the basis of these considerations it is clearly necessary to try to equate the conditions of reaction time and perceptual simultaneity tests in any experiment. A method of adjustment for simultaneity, for example, would not be comparable with any reaction-time technique because the sequential variables would be different, temporal uncertainty would not be controlled and the number of "detections" would be different. The requirements are that the task will have the same stimulus and temporal uncertainty conditions for both measures, and that variable foreperiods (essential for simple reaction time work) shall be used when both measures are taken. In addition, it is not desirable that a method of adjustment be used for simultaneity measurements since repeated exposure to a stimulus may well alter the criterion of signal presence adopted by the subject.

Experiment I

Method

Task and apparatus

It is hoped that the above conditions are met by the Pointer Test. This is an adaptation of Wundt's "complication" experiment (Wundt, 1874). The subject is required to look

at a clock face, watch a pointer revolving, and say just where the pointer was at the tim when the onset of the stimulus occurred. The independent variable here is the loudnes of the stimulus. In the reaction time task the subject is asked to release a key as soon a the stimulus has been presented, but to use the clock as an indicant of the passage of foreperiod. The same stimulus sequences with the same degree of temporal (pointe position) uncertainty can thus be used for *both* simultaneity and reaction time tests.

The "clock" was a chronoscope ("standard" timer) graduated in centiseconds with 1-sec sweep. Every tenth-of-second interval was numbered 0, 10, 20, . . . 90). Th sweep-face diameter was 8 cm and face was set centrally into a "box" 14·5 cm², and th whole chronoscope was set into a board 50 cm². The board was mounted on a table i front of the subject at a viewing distance of 60 cm. For the RT measurements a micro switch key was provided on the same table and was operated on the release-to-reac principle using the index finger of the non-preferred hand. This was to enable the subjec to record his pointer readings without undue effort. A comfortably bright level of illu-mination was used and this remained fixed during all experiments carried out with the apparatus. The lamps were arranged so that variations of illumination over the surface were minimal.

Stimuli were switched electronically into a background of random noise. They were presented after one complete sweep of the pointer. From the angle of RT, temporal uncertainty was in the range 1–2 sec, with a rectangular distribution using pointer movement as a warning signal. From the pointer test point of view, the position was in the range 0–100 (first sweep ignored).

The stimuli, like the background, were random noise and thus the subject was detecting an increment. Intensities, measured by a Dawe sound level meter coupled to the earphones were 60 dB for the background rising to 62, 63, 67, 78 dB for stimuli 1, 2, 3 and 4. They were effectively continuous. Both clock movement and stimulus were terminated after the subject had written his answer or reacted.

Subjects

The subjects were eight Naval Rating volunteers, screened audiometrically and not having a hearing loss of greater than 20 dB at any frequency on either ear.

Procedure

There were four sessions, two RT and two pointer, each lasting about 40–50 min. Four subjects performed both RT tests first and four the pointer tests first. The first session of each kind was considered practice.

(a) *RT tests.* Subjects were instructed to release the key as soon as they heard the stimulus onset. Trial-by-trial knowledge of results was given to maintain motivation.

(b) *Pointer tests.* In order to mask ties between this test and the RT test, subjects were told it was an examination of visual discrimination. They were told simply to judge where the pointer was when the stimulus came on. There is a problem with knowledge of results (KR) here in that while the variance of the answers might not be influenced very much by accurate feedback, the P.S.E. as a function of intensity certainly would, and as a result the prediction of no intensity effect would occur for purely artifactual reasons. Accordingly, KR was given as a function of *variability* on the entire test. Low *variability* was called high *accuracy* and subjects were given knowledge of this at the end of the session.

Whether tests were of reaction time or simultaneity, stimuli were presented in blocks of 42 stimuli of the same intensity with eight blank trials interspersed. The order of the four blocks in each session for each subject was determined on a latin-square basis.

Results

Only the second pointer session and the second RT session were scored.

Scoring technique

RT is described in the usual way. For the simultaneity tests "pointer errors" (sign included) are the primary data. Suppose the stimulus arrived when the clock pointer was at 58 and the subject reported 62. Then the error is +4. Had he reported 45 the error would be −13. These errors are called L values for convenience, and in describing the results we refer to mean L and mean RT for a given intensity. On a very few occasions, pointer readings were not assigned to stimuli in a given trial. Subjects on these occasions felt that they would not make a confident estimate, and since they had been requested not to guess, they simply left a blank. This, in fact, accounted for a very small proportion of trials.

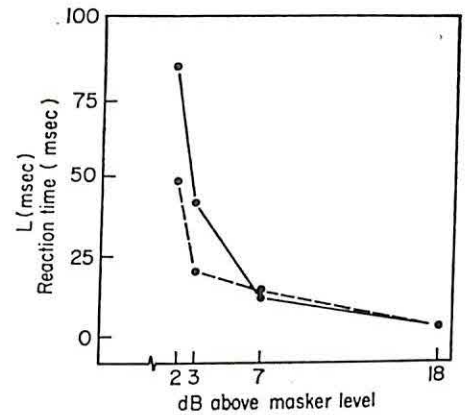

FIGURE 1. The effect of intensity upon RT and L, measuring each in separate experimental sessions. For ease of comparison RT and L have been set at zero for the loudest stimulus. —●—●—, RT; —●—●—, L.

Figure 1 shows the results and Table I gives the data for individual subjects. It is absolutely clear that intensity has a greater effect on RT than on L. Friedman's Two-way Analysis of Variance (Siegel, 1956) showed that:

(a) RT is influenced by intensity ($P < 0{\cdot}001$);

(b) L is influenced by intensity ($P < 0{\cdot}001$).

The difference between RT and L scores becomes greater as the stimulus strength is reduced ($P < 0{\cdot}001$) which confirms our expectations.

Conclusions

RT and L both depend on stimulus strength, at least when the stimuli used are in the "quiet" range. More important, they are dependent to different degrees: the mean total change in latency across the range we have used is 83 msec for RT and only 49 msec for L.

It is fairly clear that the detection process involved in judging the onset of a stimulus is not the process which determines the size of the RT-loudness effect. The result is consistent with the rating data of Sanford (1970) which suggested that there was an intensity-dependent lag between hearing a stimulus and initiating the response to it.

Throughout this stage of the experiment an effort was made to make the conditions as comparable as possible in the two tests. However, false (or premature) responses in reaction time are clearly detectable, and this alone may prevent subjects from adopting a minimal criterion. Since the subject cannot make obvious premature responses in the pointer test, it seems plausible that the

TABLE I

Individual subjects' RT and L values measured at low intensities

		Intensity (1 = softest)			
		1	2	3	4
S_1	RT	223	190	148	138
	L	79	50	40	40
S_2	RT	223	168	145	130
	L	85	40	58	29
S_3	RT	214	185	146	134
	L	149	115	90	80
S_4	RT	266	199	198	164
	L	67	21	25	19
S_5	RT	219	177	155	153
	L	79	48	39	23
S_6	RT	269	213	151	160
	L	87	57	49	38
S_7	RT	228	177	154	156
	L	80	57	40	31
S_8	RT	211	183	174	155
	L	73	69	52	46
Mean	RT	232	187	159	149
	L	87	57	49	38

criterion for the pointer test would be lower and this could explain the discrepancy between the two intensity functions. In other words, there is still a marginal possibility that both intensity functions stem from a common detection process but that despite the precautions taken the criteria adopted in each test type differ. In Experiment II, therefore, an attempt is made to control for the possibility of the adoption of different criteria in different tests by requiring the subject to make a manual reaction (RT) and a pointer judgement on the same trial. If the results found in Experiment I are duplicated then the theory of an intensity-dependent lag between detection and response initiation would be supported even more firmly.

Experiment II

Method

In this test a further eight Naval Rating volunteers produced RTs *and* simultaneity judgements in the *same* trials. Otherwise the procedure was much as before with respect to instruction, feedback, etc. Only two sessions were used, practice and test, although all

subjects had some practice at RT before this experiment. The main session consisted of 42 stimuli and eight blanks for each intensity. Intensities were again presented in blocks ordered across subjects on a latin-square basis.

Results

The pattern of results was exactly as before, with intensity affecting RT ($P < 0{\cdot}001$), L ($P < 0{\cdot}001$) and the difference between RT and L ($P < 0{\cdot}01$). The data are given in Figure 2 and Table II.

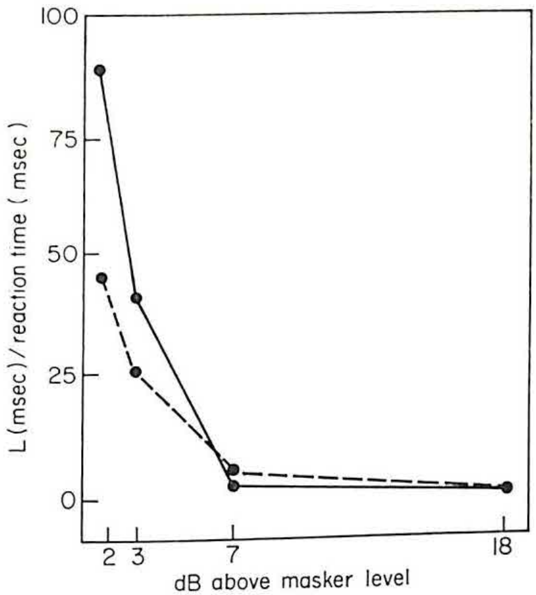

FIGURE 2. RT and L at each intensity under conditions of simultaneous measurement —●—RT -●-L.

Conclusions and General Discussion

We find confirmation of the results of the first experiment in the second experiment when key reactions and pointer judgements were taken in the *same* trials. One possible explanation of the results of the first experiment was that the intensity effect on RT was mediated through perceptual lag but that the subject could adopt a more "risky" criterion for judging the signal as present in the case of the pointer test thereby attenuating the loudness effect. This cannot be so in Experiment II however, since the subject was required to prepare for a motor reaction (RT) as well as make a pointer judgement on each trial. The difference in the sizes of the loudness effects for the two measured in Experiment II can only be due to two separate processes of perceptual detection and reaction initiation. For the auditory modality under the conditions of measurement used here, the neat dichotomy of reaction time into an intensity-dependent perceptual process and an intensity-independent reaction initiation process of the kind Roufs (1963) assumes for visual RT, seems quite inapplicable. Indeed, it is questionable whether "detection time" as measured by simultaneity techniques is a "component" of RT at all. It is just as likely that the stimulus processes in the case of perceptual

detection and reaction time are parallel and not sequential. All that is clear from the results here is that the effect of stimulus intensity on RT is not mediated by perceptual lag as measured by the pointer test.

In these investigations the stimulus intensities chosen were in the quiet-to-moderate range because a reasonable effect of intensity of RT was required. In order to generalize the findings it is desirable that the full intensity range normally encountered in RT-intensity studies be investigated, and here it may be possible to check some further issues at the same time. Namba, Yoshikawa and Kuwano

TABLE II

Individual subjects' RT and L values measured simultaneously at four intensities

		Intensity			
		1	2	3	4
S_1	RT	234	190	171	166
	L	1	1	−35	−26
S_2	RT	225	199	136	142
	L	57	68	22	0
S_3	RT	272	224	188	196
	L	92	58	33	26
S_4	RT	245	196	168	161
	L	40	−2	−5	−14
S_5	RT	283	210	163	164
	L	113	88	44	54
S_6	RT	242	197	179	173
	L	107	79	61	58
S_7	RT	250	255	200	204
	L	61	33	43	35
S_8	L	262	210	177	174
	L	68	53	29	31
Mean	RT	252	210	173	172
	L	67	47	24	21

(1968) have shown that when the range of intensities used is restricted, the difference in reaction time to the two extremes used in any range is fairly (not totally) constant. In other words there are range-context effects on reaction time. It would be possible to check whether such range effects influence the L-intensity effect as well as the RT-intensity effect. Perhaps factors influencing the magnitude of the RT-intensity effect do not effect the magnitude of the L-intensity effect. If, as the results here indicate, intensity effects on phenomenal simultaneity and reaction time are not identical, then each effect may be under the control of a different "detection" criterion.

The general findings support the view that there is an intensity-dependent lag between detecting a stimulus and initiating a response to it, thus indicating that much of the effect of auditory intensity on RT must be explained in terms of response factors rather than perceptual lag factors.

References

CHOCHOLLE, R. (1945). Variation des temps de réaction auditifs en fonction de l'intensité a diverses fréquences. *Année Psychologique*, **41–42**, 65–124.

GRICE, G. R. (1968). Stimulus intensity and response evocation. *Psychological Review*, **75**, 359–73.

JOHN, I. D. (1967*a*). Sequential effects of stimulus characteristics in a serial reaction-time task. *Australian Journal of Psychology*, **19**, 35–40.

JOHN, I. D. (1967*b*). A statistical decision theory of simple reaction time. *Australian Journal of Psychology*, **19**, 27–34.

KELLAS, G., BAUMEISTER, A. A. and WILCOX, S. J. (1969). Interactive effects of preparatory intervals, stimulus intensity and experimental design on reaction time. *Journal of Experimental Psychology*, **80**, 311–6.

KOHFIELD, D. L. (1969). Effects of the intensity of auditory and visual ready signals on simple reaction time. *Journal of Experimental Psychology*, **82**, 88–95.

NAMBA, S., YOSHIKAWA, T. and KUWANO, S. (1968). Context effects in loudness judgement. *Japanese Journal of Psychology*, **39**, 191–9.

ROUFS, J. A. J. (1963). Perception lag as a function of stimulus luminance. *Vision Research*, **3**, 81–91.

SANFORD, A. J. (1969). Unpublished Ph.D. Thesis, University of Cambridge.

SANFORD, A. J. (1970). Rating the speed of a simple reaction. *Psychonomic Science*, **21**, 333–4.

SIEGEL, S. (1956). *Non-parametric Statistics*. New York: McGraw-Hill.

WUNDT, W. (1874). *Grundzüge der Physiologischen Psychologie*. Leipzig: Engelman.

Received 4 *December* 1970

Quarterly Journal of Experimental Psychology (1971) **23**, 304–310

EFFECTS OF INFORMATION LOAD, SENSORY MODALITY, AND AGE ON PACED INSPECTION PERFORMANCE

JAMES L. FOZARD, GORDON D. CARR AND GEORGE A. TALLAND
Veterans Administration Outpatient Clinic and Harvard Medical School, Boston, Massachusetts

AND

DONALD E. ERWIN
Massachusetts General Hospital, Boston, Massachusetts

The effects on paced inspection performance of amount of stimulus information, presentation of information visually, as opposed to visually and auditorily, and age were investigated. In comparison to performance with one sequence of visual information, correct identifications of signals declined and false alarms increased when two sequences of visual information were monitored. Increasing the number of classes of signals in the two sequences did not significantly affect performance. However, when one of the two sequences was presented visually and the other auditorily, performance improved. There was a tendency for performance to be lower among older subjects in most experimental conditions.

Introduction

The experiment required the subject to monitor a sequentially presented series consisting of letters and digits and to report the occurrence of a signal, here defined as a sequence of three consecutive letters or three consecutive digits. Thus, the task required continuous registration and discarding of information in order to identify signals correctly. Men of varying ages performed the task under four experimental conditions. In the first, two types of signals were presented in one visual sequence. In a second, two types of signals were presented in each of two visual sequences. The third condition contained four rather than two types of signals in each of the two sequences. The fourth condition was also similar to the second with the exception that one sequence was presented visually and the other auditorily. Decision and signal rates were held constant across all conditions.

Results from the four conditions bear on three questions: (1) What is the effect on performance of requiring the subject to attend to two independent sequences of information as opposed to only one sequence? (2) Is it easier to identify signals in two sequences when they are presented in two sensory modalities rather than one? (3) What is the effect on performance of increasing the number of types of signals in two independent sequences? In each case, the data were examined to determine whether any effects of the experimental treatments interact with age.

The present experiment would be classified by Postman (1964) as "continuous instrumental responding". Since there was more than one type of signal used, the experiment has some bearing on Postman's hypothesis that changing the type of successive signals (i.e. from letters to digits or vice versa) reduces interitem interference, thereby improving performance. Postman argues that because of stimulus generalization, a subject may be more likely not to discriminate critical from non-critical stimuli when successive stimuli are the same than when they are different. Some light was shed on this question by comparing the number of correct identifications of the second of two successive signals of the same type with successive signals of different types.

The probability of the detection of a given signal has been related to the length of time between it and the previous signal in a task similar to the present one (McGrath and Harabedian, 1963). Thus, a final question which may be asked of the data is whether any such relationship is modified by the present experimental treatments.

Method

Subjects

The sample consisted of 46 men whose ages ranged from 25–69. All were participants in a longitudinal study of the ageing process conducted by the Boston Veterans Administration Outpatient Clinic (Bell, Rose and Damon, 1966). By various medical criteria, subjects were in above average health at the time of entry into the study, and with the exception of a few older subjects, all were veterans (Bell, Rose and Damon, 1966). Educational backgrounds varied from several years of high school to some study at the graduate level, and socio-economic class from Warner Levels one to seven (Warner, Meeker and Eells, 1960). Typical occupations were police officer, computer operator and insurance broker. For purposes of analysis, the sample was divided into three age groups: 17 young subjects aged 25–41; 30 middle-aged subjects aged 42–55; and 9 old subjects aged 56–70 years. The range of variation in education and socio-economic status was similar in the three age groups.

Procedure

Each subject served in four experimental conditions. The order of presentation of conditions was varied across subjects, each possible order being presented to at least one subject. In the visual conditions, the subject was presented with a combined sequence of digits and letters appearing one at a time in the aperture of a Gerbrand's Memory Drum. In the audio-visual condition, the memory drum was used to present one sequence of letters and digits, but concurrent with the visual presentation, a second list was presented auditorily by a tape recorder. In all conditions, the subject's task was to call out the last stimulus of any successive series of three stimuli of a particular class.

Conditions

The four experimental conditions are described below:

*V*12: A single sequence of letters and digits were presented at the rate of one per second. There were two types of signals: series of three successive letters or three successive digits.

*V*22: Two independent sequences of visual symbols were presented at the rate of one pair (i.e. one symbol from each sequence) per 2 sec. There were two types of signals as in V12. These had an equal likelihood of occurrence in either list. Simultaneous signals in both sequences did not occur.

V24: Two sequences of visual symbols were presented as in V22. There were four types of signals: series of three successive upper or or lower case letters, or, odd or even digits. All had an approximately equal likelihood of occurrence in either sequence, and were also presented at the rate of one pair per two sec.

AV22: One sequence of visual and one sequence of auditory symbols were presented at the rate of one symbol per 2 sec. Signals were arranged the same as in V22.

In each condition a total of 360 symbols was presented over a period of 6 min. In Conditions V12, V22 and AV22, there were 24 signals, while in V24 there were 25 signals. The intervals between signals ranged from 0 to 30 sec, the modal interval being 7 sec. Thus, in V12, where the memory drum advanced at the rate of one per sec, a 10-sec intersignal interval consisted of the presentation of ten non-signal symbols. In other conditions, where the symbols were presented at the rate of one pair per two sec, 10-sec intersignal interval also meant the presentation of ten non-signals, five in each sequence. Correct reports of the last symbol of a signal triad were recorded as correct detections. Incorrect reports of such symbols and reports of all other symbols were recorded as false alarms.

In summary, the comparison of conditions V12, AV22 and V24 to condition V22 tests the effects on performance when the subject is required to: (1) monitor only one sequence of visual presentations (V12) rather than two (V22); (2) monitor audio and visual sequences of presentations (AV22) rather than merely two visual sequences (V22); and (3) monitor four (V24) as opposed to two (V22) types of stimuli in each sequence of stimuli—although the amount and rate of information presented remains the same.

Results

Preliminary analyses revealed no differences in performance associated either with classes of stimuli, i.e. letters vs. digits, or, in the visual conditions, with the left vs. right sequences; no practice effects for either letter or digit signals within or between conditions; and indicated that performance did not vary according to the order of presentation of the four conditions. There were no systematic differences in performance associated with education or socio-economic status. Therefore the percentage of correct signal identifications were averaged across letter and digit signals as is shown in Table I.

TABLE I

Per cent of correct identifications of signals in the four experimental conditions by subjects in three age groups

Age Groups	Condition V12	V22	AV22	V24
25–41	85·9	42·8	54·9	40·4
42–55	86·7	42·7	42·7	39·1
56–70	81·0	29·9	51·3	30·9

Number of sequences monitored

In all age groups the most consistent difference in performance among the four conditions was between those involving one and two sequences of information. Only the data from Conditions V12 and V22 were compared in a statistical analysis because the other conditions involving the two sequences varied from Condition V12 in either the number of different types of signals that needed to be monitored

(V24) or, in the sensory modality used (AV22). Results showed that only the main effect of Condition was statistically significant, $F(1, 43) = 220{\cdot}32$, $P < 0{\cdot}001$.

Two sense modalities

Table I shows that dividing the two sequences between two sensory modalities results in better performance than presenting both sequences in the same modality. Results of an analysis of variance performed on the data from Conditions V22 and AV22 showed that the difference in performance in the two conditions was significant, $F(1,43) = 6{\cdot}54$, $P < 0{\cdot}01$. Table I suggests an age-related decline in performance in Condition V22 but not in AV22, implying that in comparison to younger subjects, older subjects are aided relatively more by having the same amount of information presented in two rather than one sensory modalities. However, this interaction fell slightly short of statistical significance, $F(2,43) = 2{\cdot}62$.

In Condition V22 there was a tendency for detection to decline with longer intervals while in AV22, detection improved with longer intervals. The difference in these trends was not significant; they appeared to be associated with differences in the frequency distributions of inter-signal intervals in the two conditions. These distributions were positively skewed in the visual conditions and nearly rectangular in the audio-visual condition. McGrath and Harabedian (1963) have obtained similar results with vigilance tasks, finding improved detection as a function of interval associated with rectangular distributions and somewhat decreased detection as a function of interval associated with a positively skewed distribution. (There were no effects of inter-signal interval size on performance in Conditions V24 and V12.)

Number of signal types

Table I shows that, regardless of the subject's age, it was no more difficult to monitor four rather than two types of signal. Results of an analysis of variance performed on the data from Conditions V22 and V24 showed that neither the effects of Condition, Age, nor their interaction were statistically significant.

In Conditions V22 and V24 auxilliary analyses were performed to determine the effect on correct detection of alternating or repeating the class of symbol in successive signals. In V24, "same" meant either the same case and class (e.g. critical series of three small letters followed by critical series of three small letters) or same class only (e.g. critical series of three small letters followed by a critical series of three capital letters). In no comparisons were there any consistent differences associated with the variable.

Auxilliary analyses indicated that correct and incorrect identifications of a signal in one sequence were not affected by variations in the number of symbols of the same class that appeared either in the other sequence or shortly before or after a critical series in the same sequence. (This conclusion rests heavily on analyses of Condition V22; there were comparatively few relevant instances in Condition V24.) Apparently, cross-sequence repetitions of stimuli of the same class as the signal had little effect on performance.

False alarms

The average incidence of false alarms was very small; in no condition did it exceed 1%. There were fewer false alarms in Condition V12 (0·2%) than in Conditions AV22 (0·6%), V24 (0·6%), or V22 (0·8%). The differences in the number of false alarms in the four conditions was statistically significant, $F(3,43) = 10 \cdot 24$, $P < 0 \cdot 01$. There were no significant age-related differences in the number of false alarms, nor was there any evidence that the occurrence of false alarms was related to particular symbol sequences either within a sequence or across sequences.

The small number of false alarms precluded a conventional signal detection analysis. Therefore, in order to obtain a measure of overall performance which would include both correct detections and false alarms, a "percentage correct decisions" score was calculated for each subject under each task condition (Blosser, 1965). This score was equal to the sum of the weighted proportions of correctly detected signals and correctly rejected non-signals. It thus represented the proportion of all task stimuli on which the subject made the correct decision as indicated by his appropriately calling out the third character of a triad or witholding a response.

The rank order of the averages of the calculated scores was identical to that of the means in Table I. The same statistical analyses described for the percentage of correct identifications were performed on the percentage correct decision scores. The results were the same in all cases as those reported earlier.

Discussion

The results indicate that presenting visual information in two sequences rather than in one impairs performance. Dividing the two sequences of information between two sensory modalities, however, results in an improvement in both correct detection of signals and overall performance. When two sequences are presented in the same sensory modality, increasing the number of types of signals in each sequence does not significantly affect performance.

In the two-sequence conditions, storage load per presentation is greater, and consequently performance is depressed (Reid, Lloyd, Brackett, and Hawkins, 1961). For example, in Condition V12, each letter or digit represents first-order information—stimulus one equals A, or 2, etc. However, in Condition V22, each stimulus has second order informational content as well—stimulus one equals *left* "A", or *right* "2", etc. As those conditions where stimuli have second-order informational content result in impaired performance, these results are thus compatible with Broadbent's (1958) conclusion that the information capacity of the human is limited, and increasing the number of sequences impairs performance.

One would not expect confusions between stimuli in the two sequences at the relatively slow presentation rates employed (Moray, 1970). And indeed, in none of the two-sequence visual conditions was there any evidence of differences between correct or incorrect identifications of signals in the left and right sequences. Nor was there any evidence of confusions between items across the two sequences.

Performance was significantly better in Condition AV22 than in V22. This

suggests that eye-ear coding of information was more efficient than "Left-Right" coding. However, coding may have been further assisted in AV22 by the asynchrony between the visual and auditory sequences. This asynchrony would be expected to render the two sources of information even more discriminable, leading to still more efficient coding than was possible in V22.

Altering Condition V22 by doubling the number of types of signals per channel (V24) produced no significant impairment of performance. The false alarm rate for Condition V24 did not exceed that of V22, which would have happened if stimulus generalization had occurred. Thus, increased inter-item interference (Postman, 1964) does not seem to be occurring in Condition V24. In some respects, the present tasks resemble the sequence of events in short-term memory studies where changes in class of stimuli in a series of trials have been found to improve recall, i.e. effecting a release from proactive inhibition (Wickens, 1970). In this case, slightly improved performance in Condition V24 following longer intersignal intervals (10–15 sec) may be due to the increased number of types of stimuli resulting in a release from proactive inhibition.

There were consistent but insignificant trends towards poorer performance in the older subjects under all experimental conditions. The presentation rates used here were relatively slow. Probably faster presentation rates would have accentuated these apparent age-related performance trends (Davies, 1968). At the two sec rate used in the two-sequence conditions, some rehearsal may have been possible and could have been relatively more beneficial to the older subjects.

In Condition AV22, the most interesting age-related effect was the lesser decline in performance than that found in V22. If the older subject's information capacity is lower or less efficient than the younger subject's, it is plausible that the older subject would benefit more by the sequential, more discriminable (and less "loaded") presentation of information in two sensory modalities than is the case in Condition V22 (Reid *et al.*, 1961). The relatively greater benefits experienced by older subjects when aided by dividing the information into two sensory modalities is an interesting effect requiring further research. The practical applications of this finding lie in the field of human factors engineering where the capacities of older individuals need to be considered. A more exacting description of this effect is warranted.

These experiments were conducted at the Stanley Cobb Laboratories for Psychiatric Research at the Massachusetts General Hospital. They were supported by Grant HD00340 from the National Institute of Child Health and Human Development, the Veterans Administration Outpatient Clinic, Boston, Mass., U.S.A., and the Council for Tobacco Research, U.S.A. Thanks are due to Misses Anne Pope and Dena Scher who collected and organized the data for analysis.

References

Bell, B., Rose, C. L. and Damon, A. (1966). The Veterans Administration longitudinal study of healthy aging. *Gerontologist*, **6**, 179–84.

Blosser, A. B. (1965). *A Performance-oriented Approach to Detection: Tables for Detection, Discrimination and Decision Theory*. Houston Texas: Tracor, Inc.

Broadbent, D. E. (1958). *Perception and Communication*. Oxford: Pergamon Press.

DAVIES, D. R. (1968). Age differences in paced inspection tasks. In Talland G. A. (Ed.) *Human Aging and Behavior*. New York: Academic Press.

MCGRATH, J. J., AND HARABEDIAN, A. (1963). Signal detection as a function of intersignal-interval duration. In BUCKNER, P. N. and MCGRATH, J. J. (Eds.) *Vigilance: A Symposium*. New York: McGraw-Hill.

MORAY, N. (1970). *Attention: Selective Processes in Vision and Hearing*. New York: Academic Press.

POSTMAN, L. (1964). Short-term memory and incidental learning. In MELTON, A. W. (Ed.) *Categories of Human Learning*. New York: Academic Press.

REID, L. S., LLOYD, K. E., BRACKETT, H. R. AND HAWKINS, W. F. (1961). Short-term retention as a function of average storage load and average load reduction. *Journal of Experimental Psychology*, **62**, 518–22.

WARNER, W. L., MEEKER, M. AND EELLS, K. (1960). *Social Class in America: The Evaluation of Status*. New York: Harper.

WICKENS, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, **77** (1), 1–15.

Received 7 December 1970

Quarterly Journal of Experimental Psychology (1971) 23, 311–315

EXTENSION CUES IN OPEN AND CLOSED FIGURES

NORMAN WORRALL AND DIANE FIRTH
Department of Psychology and Social Studies, Hatfield Polytechnic

Subjects were shown line drawings of figures differing in degree of closure, and asked what the drawings most looked like. Responses were dichotomized into "upright" (e.g. beaker) and "extended" (e.g. runway). Although, overall, about twice as many upright as extended responses were given, relatively open figures yielded about twice as many extended responses as did relatively closed figures. Converging lines as such appeared an unexpectedly weak cue for extension, and horizontal discontinuity in a drawing seemed more suggestive of extension than did vertical discontinuity. Reclassifying the responses as two- or three-dimensional showed that although two-dimensional responses were less common overall, they were three times as common in relatively closed drawings. Predictions are made relating the findings to standard visual illusions.

Introduction

The reasons why some line drawings suggest extended depth while others suggest flat figures have not been established. This paper reports an experiment investigating what kinds of cues cause the observer to report that certain line drawings are "upright" as opposed to "extended": what decides, for example, whether a general triangular figure is seen as a tent (upright) or as a road going into the distance (extended). Monocular discrimination in the laboratory between a square frame viewed obliquely and the upright figure which would fill the same solid visual angle is notoriously difficult, and we have made an analogous enquiry into how readily extension or uprightness is ascribed to different types of line drawings, particularly those employing the "classical" cue for extension—a pair of converging lines.

Casual observation suggests that the extension or uprightness of line drawings is related in an important way to the degree of closure in the figure. For example, in Figure 1, drawing 2(b) would seem to be most readily seen as the conventional representation of a road going away from the observer into the distance. However, when the converging lines are closed to form an apex [drawing 2(a)] the figure changes palpably to represent something upright. It was felt generally that the extent to which a line drawing might be characterized as "open" or "closed" could have significant bearing on the perception of extended depth in such a figure. The terms "open" and "closed" are used here to refer to the number of free ends in a figure; in two figures containing the same number of lines, the figure with the greater number of free ends is regarded as "open" relative to its fellow (see Fig. 1).

Method

India ink drawings were presented on 18 in. square white card to small groups of subjects. "Closed" drawings had converging lines 15 in. long, while the "open" versions each had a 3 in. slice removed at the apex. Lines were 4 mm thick, and had a constant angle of convergence of 60 degrees. In Figure 1, the left-hand drawing of each pair is closed relative to the right-hand drawing which introduces one stage of discontinuity into the figure. Thus, drawing 1(a) is a completely closed triangle which is opened in 1(b) by removing the apex. In 2(a) the triangle is opened by removing the base, while 2(b) introduces a further stage of discontinuity by, in addition, removing the apex. Other forms of closure could have been used; the drawings in Fig. 1, while not exhaustive, are intended to be representative of the general problem.

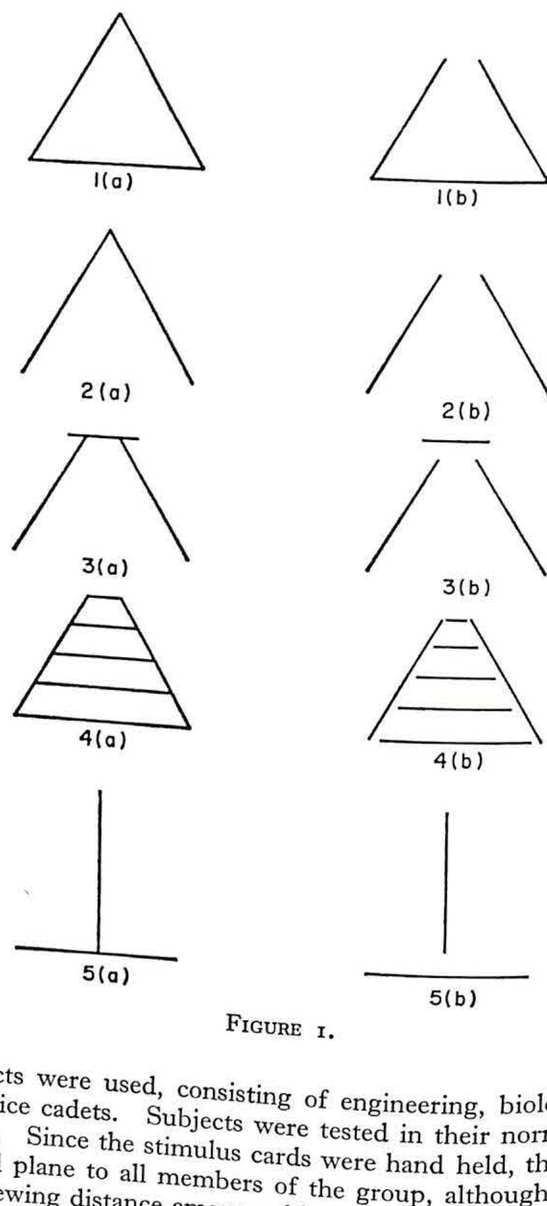

FIGURE 1.

Seventy-five subjects were used, consisting of engineering, biology and social studies students and also police cadets. Subjects were tested in their normal seminar groups of from 4 to 12 persons. Since the stimulus cards were hand held, they could be presented in the frontal–parallel plane to all members of the group, although there was necessarily some variability in viewing distance among subjects. Each figure was exposed for 10 sec, with a pause of approximately 20 sec between successive presentations. Subjects were

asked to answer the question "what does the drawing most look like to you?". Order of presentation was randomized for each group.

Three judges scored the data. One was a journalist (J), one a sociologist (S) and the third a psychologist (P), the present senior author. For each drawing, judges were asked first to read through all responses given to a particular pair of drawings so as to obtain general impressions, then to work through a second time scoring each response as "upright" or "extended". For each drawing, the proportion of extended responses was calculated, based on scorable items. Not more than six responses out of a possible total of 75 were lost for any drawing through inability to score or because no response was given. Clearly, responses such as "road going under a bridge" [3(a)] or "worms on a conveyor belt" [4(b)] would be scored as extended, while "table" [again 3(a)] or "behive" [again 4(b)] would be scored as upright. However, each judge would have his own criteria, especially when more difficult distinctions had to be made, for example between "step-ladder" and "staircase". The same scoring criteria were applied to all drawings, and the most difficult responses were not scored—for example, "billiard frame" [1(a)]. No judge was unable to score more than two responses for any one drawing.

Results

The proportion of extended responses, as scored by each judge for each figure, is shown in Table I. Although only judge P was aware of the drawings that had given rise to each set of responses, there is surprisingly little variability between judges: the largest discrepancy is for drawing 5(b). When the differences within each pair of drawings are tested using the binomial test, the difference is significant ($P < 0{\cdot}01$) for each pair, except pair 3. This is true for the ratings of all three judges.

TABLE I

Figure	Per cent extended Judge P	Judge S	Judge J	Mean	Per cent three-dimensionality
1(a)	5	1	3	3	14
1(b)	12	12	14	13	58
2(a)	9	6	6	7	62
2(b)	61	63	61	62	73
3(a)	34	41	33	36	78
3(b)	40	38	35	38	66
4(a)	45	47	38	43	85
4(b)	81	75	83	80	100
5(a)	10	4	13	9	57
5(b)	21	48	41	37	63
Mean per cent for closed figures	21	20	19	20	59
Mean per cent for open figures	43	47	47	46	72

Of all the responses scored in the experiment, only about a third were extended. This is in spite of the fact that most figures contain converging lines, commonly regarded as an important cue for extension. Of the extended responses that were

given, two-thirds were to the open drawings, and only one-third to the closed drawings.

It can be seen that the closed triangle [1(a)] yields very few extended responses indeed, but when the apex is opened [1(b)] there is a significant increase in extended responses. Opening the triangle at the base [2(a)] also increases extended responses, but the really large increase occurs at two stages of discontinuity from the closed triangle [2(b)]. Partly closing drawing 2(b) with a horizontal correspondingly halves the number of extended responses [3(a)]. Drawing 3(b) does not continue the pattern, since opening the horizontal would now be expected to increase the extended responses. There may be a special problem here in that both 3(a) and 3(b) tended to evoke the stereotype response, "road under a bridge" or some variant of this. Nor is there any way of knowing whether the open version [3(b)] may have been a more "convincing" representation of extension than drawing 3(a). However, the pattern is picked up again in pair 4, where 4(a) closes the converging lines with a series of horizontal ties. Even though there is now in 4(a) a dual cue for extension—converging lines plus horizontals of decreasing length—the extended responses are nevertheless fewer than those evoked by simple, open converging lines [2(b)]. Only by opening the figure, as in drawing 4(b), do the extension-suggesting properties of these combined cues become fully operative, and in fact drawing 4(b) yields the largest proportion of extended responses for any figure used in the study—80%. Finally, 5(a) and 5(b) show what happens when a vertical is separated from a horizontal—extended responses at least double. For this pair, judges S and J were at variance with judge P. To give an example, the former scored "flagpole" as upright, but "flagpole on a lawn" as extended, whereas judge P, being aware of the figures, required an answer more closely implying distance, such as "flagpole across a lawn".

It is also possible to classify the responses on the more orthodox dimension of two- or three-dimensionality. Thus, all extended responses are automatically three-dimensional, together with "solid object" responses from the upright category such as "tent" or "rocket". Responses referring to geometrical figures, diagrams, signs or symbols were classified as two-dimensional, for example, "part of a triangle", "pi", "operational amplifier". Since two- or three-dimensionality is more a matter of fact and less a matter of judgement, only judge P scored the responses on this dimension. It can be seen from Table I that for all pairs, except again the anomalous pair 3, the open figures have a somewhat greater proportion of extended responses than have the closed figures. Overall, 59% of the responses to closed figures were three-dimensional, while the value for open figures was 72%. Generally, therefore, the differences between open and closed figures on this dimension are less impressive than those found on the extended-upright dimension.

Discussion

It seems that the extent to which a drawing can be characterized as open or closed is important in determining whether extension will be perceived in that drawing. Continuous or tied lines increase apparent uprightness, while discontinuous or free lines increase impressions of extension. For the present data,

only in the case of two figures does the proportion of extended responses rise above one half, drawings 2(b) and 4(b), in spite of the pronounced presence in most drawings of converging lines. Thus, converging lines are not a strong cue for extension over distance. In fact, somewhat paradoxically, the present data suggest they are a strong cue for uprightness rather than extension. To operate as effective cues for "linear perspective" converging lines must be discontinuous with each other and with other lines in the figure. If they are at all tied, at apex [drawing 2(a)], base [1(b)], both apex and base [1(a)], or even middle [4(a)], the effect is to reduce the probability of an extended response. Especially convincing is the case of pair 4, where two cues independently regarded as being strong indicators of perspective, are as a result of *themselves* being tied together, made ineffective relative to the case of only *one* of these cues operating but in open form [2(b)].

At a simple level of interpretation, discontinuity in lines does seem to allow the possibility for space to be interpolated. One might expect relatively closed figures to have better object-stability in a real-life sense than relatively open figures, which would be more unstable. In the case of pair 5, for example, where the open figure shows vertical discontinuity, this would seem quite reasonably to increase the impression of a perspective extending before the observer, and is a device commonly used by the sketch artist. Indeed, the phenomenon would be relatively straightforward if it had been found that only vertical discontinuity in a drawing gives rise to impressions of extension. However, of the more remarkable contrasts in the data are for pairs where the only difference between the figures is in terms of horizontal, not vertical, discontinuity (pairs 1, 2 and 4). Further, in the admittedly anomalous pair 3, vertical discontinuity makes no difference to the proportion of extended responses.

While it is hoped that development of the approach suggested in this study may help formulate rules about what kinds of retinal stimulation patterns lead, for example, to perception of solid or plane figures, the present findings already suggest certain predictions with regard to the visual illusions. For example, if the Ponzo illusion depends on perspective cues, the illusion will be less effective when either the base or the apex of the containing figure is closed. It will be least effective when the figure is maximally closed (triangle) and most effective when the figure is maximally open (open converging lines). Again, if open figures in general have stronger depth and extension features than closed figures, and if the "inappropriate constancy scaling" account of the Müller–Lyer illusion is correct (Gregory, 1966) then in versions of the illusion where the fins and shaft are discontinuous the illusion should be stronger. Inasmuch as the facts do not support this prediction, the inappropriate constancy scaling argument is weakened.

Reference

Gregory, R. L. (1966). *Eye and Brain.* London: Weidenfeld and Nicolson.

Received 21 *January* 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 316–323

AGE DIFFERENCES IN RECOGNITION MEMORY

FERGUS I. M. CRAIK

Birkbeck College, University of London

It is generally believed that short-term memory (STM) performance decreases from maturity to old age, but the present paper questions the generality of this conclusion. The first experiment reported is a short-term recognition probe study—no significant age differences were found in either acquisition or the rate of forgetting. It is suggested that age decrements in STM are limited to situations where attention is divided at input or where the material must be manipulated during storage. Experiment II confirmed previous findings of minimal age losses in secondary memory word recognition. The results of this second study support the conclusion that there is poorer accessibility to stored verbal material with advancing age.

Introduction

Since the publication of Welford's (1958) influential book, it has been widely accepted that performance on short-term memory (STM) tasks declines from maturity to old age. Apart from the evidence reported by Welford, this conclusion is supported by the work of Broadbent and Heron (1962), Inglis and Caird (1963), Kirchner (1958) and Talland (1965).

There is evidence which opposes this conclusion, however. Studies of memory span have generally shown little or no age decrement, (Bromley, 1958; Talland, 1965) and the last few words in free recall lists are also recalled as well by older subjects as they are by a young group (Baddeley, unpublished data; Craik, 1968). Both memory span and recall from the recency positions in free recall could be regarded as estimates of "primary memory", thus the question arises as to why some studies have shown poorer short-term retention performance with advancing age, while others have not. When a closer look is taken at the studies which have shown age decrements in STM, the common factor seems to be that the tasks all required the subjects to divide their attention between two processes; either two sources of input (Inglis and Caird, 1963), perception and storage (Broadbent and Heron, 1962) or storage and response (Kirchner, 1958; Taub, 1968). This observation led Welford (1958) to suggest that STM was more vulnerable to interference in older subjects.

The principle that STM is particularly fragile in the elderly and easily disrupted by interference or distraction is an attractive one and does appear to integrate a variety of findings. It would seem that the idea could be directly tested by the Peterson procedure in which items are presented to the subject for short-term retention and the interval is filled with a distracting task such as counting backwards. The proportion of items remembered drops rapidly over the first few

seconds of the retention interval and reaches an asymptote after only 15–20 sec. The prediction from Welford's hypothesis is that older subjects should show a faster rate of forgetting but, surprisingly, the experiments which have been carried out have shown no such effect. Studies by Kriauciunas (1968), and Talland (1967) gave the result that, while the overall level of retention was somewhat lower in older age groups, the rate of forgetting was no faster. Further studies by Keevil-Rogers and Schnore (1969) and by Taub and Walker (1970) also failed to show that STM was more vulnerable to interference in older subjects. In all these studies, attention was not divided when the items to be remembered were presented and it may be that older subjects are only penalized when attention is divided at input. In this case the age decrement would reflect a failure of acquisition, however, and not a failure of STM as such.

The present paper reports two short studies which examined age differences in short-term recognition, both in primary memory (Experiment I) and secondary memory (Experiment II). Signal detection analyses were carried out since it seemed likely that the more sensitive techniques of signal detection theory (SDT) would throw further light on age differences in acquisition and forgetting. Attention was not divided at input in either study, thus a failure to find age differences in memory would be further indirect evidence that division of attention is crucial to the finding of an "age decrement in STM".

Wickelgren and Norman (1966) described a short-term recognition memory technique which yields independent measures of acquisition and forgetting. A series of items is presented and followed immediately by a probe item; the subject's task is to decide whether or not the probe was a member of the series. The probe is in the list on, say, half of the trials and in these cases, its serial position is varied. Thus the "hit-rate" and "false alarm rate" for different serial positions may be obtained and d' calculated from these values. Wickelgren and Norman argue that d' is an index of trace strength and show that the value of d' decreases exponentially as the probe item occurs further back into the presentation list. Thus a graph of log d' against serial position (plotted from the last item back into the list) is a straight line whose intercept on the ordinate is a measure of acquisition strength and whose slope is a measure of the rate of forgetting in STM. The first experiment reported in the present paper is an application of this technique to age differences in short-term recognition memory, the questions at issue are whether an age decrement is found when attention is not divided at input and if a decrement does appear, whether it is attributable to poorer acquisition or more rapid forgetting. "Interference" in the present paradigm is represented by the need to perceive and store succeeding items in the presentation list.

Experiment I

Method

Each trial consisted of seven two-digit numbers presented auditorily to the subject at the rate of two digits per 1·5 sec. The last pair of digits was followed after 1·5 sec by a tone and this signal was followed after a further 1·5 sec by a test pair of digits. The subject's task was to decide whether the test pair had occurred in the preceding series. The digits were recorded by a male speaker and were pronounced as single digits—thus "13, 49, 04" was

spoken "one–three, four–nine, oh–four". On half of the trials the test pair had ocurred in the presentation series and on these occasions the critical pair was located equally often in serial positions 2–6. The first and last serial position were not tested in order to avoid primacy and ceiling effects respectively (Wickelgren and Norman, 1966).

After the task was explained to the subject he was given five practice trials followed by 50 test trials. Of the 50 trials, 25 had "positive probes" in the sense that the test pair had already been presented, and five such probes were located at each of serial positions 2, 3, 4, 5 and 6. Four sequences of five practice trials plus 50 test trials were recorded, using different digit pairs and a different pattern of positive and negative probes. Within each age group of 16 subjects, each sequence was given to four subjects.

Subjects were tested singly or in small groups and were instructed to listen passively to the presentation series on each trial, not to actively rehearse the digits. After each trial, they responded "yes" or "no" by writing on a prepared answer sheet. The digits were played to the subject at a comfortable volume through a loudspeaker.

Three age groups of subjects performed the experiment: Young (range = 19–30 years, mean age = 23·1 years), Middle (range = 35–59, mean = 46·7) and Old (range = 61–85, mean = 76·1). There were 16 subjects in each group. The age groups were matched on Mill Hill Vocabulary (MHV) score since verbal intelligence declines little with age (Welford, 1958). Mean MHV scores were 36·0, 36·9 and 35·9 for the Young, Middle and Old groups respectively.

Results

Table 1 shows that the proportion of digit pairs correctly recognized declined as the position of the probed pair was moved towards the beginning of the list. The table also shows that age differences were slight at all serial positions. An analysis of variance on the recognition scores yielded a significant effect of serial position ($F = 31{\cdot}75$, df 4, 180, $P < 0{\cdot}01$) but no effect of age or of an age × serial position interaction ($F < 1$ in both cases). The conclusion drawn from the raw data analysis is therefore that age does not affect performance on short-term recognition tasks of this type.

TABLE I

Proportion of correct recognitions at each serial position for three age groups

	Serial position				
	6	5	4	3	2
Young	0·92	0·82	0·72	0·56	0·56
Middle	0·89	0·81	0·65	0·55	0·62
Old	0·89	0·80	0·65	0·56	0·52

However, Tune (1966) found evidence from a vigilance task that older subjects maintained their detection rate at the level of younger subjects by making more false positive errors and it is thus possible that the older subjects in the present experiment behaved in a similar fashion. A signal detection analysis (which takes false positives into account) was thus carried out on the data. Hit rates were obtained for each serial position and the simplifying assumption made that the false alarm rate remained constant over all serial positions. This assumption is unsatisfactory but there is no simple way of allocating false positives to specific serial positions. Figure 1 shows values of log d' plotted against serial position. Wickelgren and

Norman (1966) found that log d' was a linear function of serial position provided that the first item in the list was ignored. First items consistently fell above the best-fitting straight line and the authors argued that this superiority was attributable to primacy. Figure 1 shows that straight lines fit the present data quite well but the *second* items in the list gave better performance than that predicted from the later serial positions. It is possible that in the present study instructions not to rehearse were not given so forcefully and that subjects did in fact rehearse the first one or two items. Accordingly, best-fit straight lines were calculated by the least squares method for serial positions 3–6 and are shown in Figure 1. The figure shows a slight age decrement in the height of the lines (reflecting poorer acquisition by Wickelgren and Norman's argument) but no steeper slope with age (thus no faster forgetting in the elderly). These conclusions are strengthened by the results of an analysis of variance on the log d' data from serial positions 3, 4, 5 and 6. The effect of serial position was significant ($F = 29{\cdot}87$, df 3, 135, $P < 0{\cdot}001$) but neither age nor the age × serial position interaction approached significance.

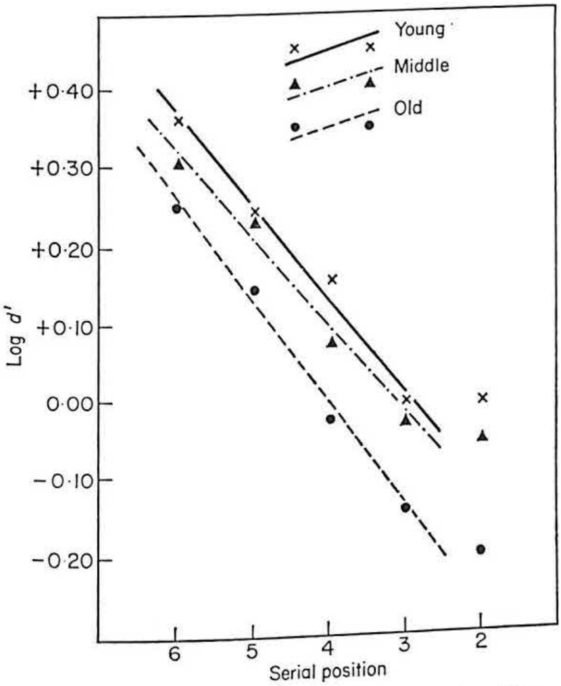

FIGURE 1. Log d' scores as a function of age and serial position.

The assumption that the false alarm rate is invariant with serial position in a short-term task of this nature is somewhat unlikely. It is possible to make another assumption—that the criterion is invariant with serial position—and calculate values of d' on this basis. This was done by first pooling the data for each subject and obtaining an overall value of β for that subject. Values of d' were then calculated for each serial position from the relevant hit rate and the common β. This procedure yielded group means which were very similar to those shown in Figure 1. An analysis of variance also gave a significant effect

of serial position ($F = 28 \cdot 23$, df 3, 135, $P < 0 \cdot 001$) but no significant effect of age or of the age $\times$ serial position interaction.

Thus, analysis of the correct recognition scores and two signal detection analyses all yielded the same result: that recognition is poorer for earlier items but that age has no reliable effect on this ability. The SDT analyses are preferable to the analysis of raw scores since they take errors of commission into account, but both SDT analyses are based on unsatisfactory assumptions. The assumption that false alarm rate does not change with serial position is not very plausible and there is evidence from a paired-associate recall study (Murdock, 1966) that the criterion also changes with serial position. However, both SDT analyses confirmed Wickelgren and Norman's (1966) finding that log d' decreases linearly with serial position and, on the basis of their arguments, it is concluded that in the present experiment neither acquisition nor rate of forgetting was impaired in the older groups.

Experiment II

The second experiment studied age differences in slightly longer term recognition memory. Schonfield and Robertson (1966) found large age decrements in recall of words but no significant age differences in recognition. They suggested that older subjects show no losses in acquisition and storage but are much less efficient at retrieval of the material in store. This argument follows from the notion that while recall involves a "search" of memory, the search process is largely bypassed in the case of recognition. Thus, the words are just as available but less accessible for older subjects. This conclusion also follows from a study by Laurence (1967) who found an age decrement in non-cued recall but none in cued recall. McNulty and Caird (1966) argued against the accessibility hypothesis, however. They made the point that acquisition and storage may actually be poorer in older subjects but that the partial information available may be sufficient to yield good recognition scores although it is not sufficient for adequate recall.

The present study examined age differences in recognition using a signal detection analysis. The technique offers two advantages over traditional methods in the present context. First, as mentioned previously, possible age differences in criterion are taken into account. Second, since the d' measures yielded by SDT may be more sensitive than traditional measures, they should demonstrate the age decrement in trace strength suggested by McNulty and Caird.

Method

A list of 25 common two-syllable nouns was read to subjects at a 1·5-sec rate. Subjects were told to attend to the words as they would be asked to recognize them later. Immediately following the presentation list, a list of 50 words was read to the subjects at a rate of one word every 5 sec. This second list comprised the 25 original words plus 25 distractors of a similar type, mixed at random. After hearing each word, subjects wrote their recognition responses "yes" or "no" followed by a confidence rating 3, 2 or 1 for "certain", "fairly sure" or "guess" respectively. The net effect of this procedure was to yield a six-point confidence scale running from "certain test word was present" to "certain test word was absent".

Ten subjects in each of 3 age groups performed the task. The groups were: Young (age range = 19–28; mean age = 25·0; mean MHV = 35·6), Middle (age range = 35–47; mean age = 41·4; mean MHV = 35·9) and Old (age range = 55–84; mean age = 71·3; mean MHV = 35·6).

Results

The data were pooled for each age group and are shown as memory operating characteristic (MOC) curves in Figure 2. Values of d' may be obtained from this graph by subtracting the false alarm rate from the hit rate (both measured in normal deviate units) at the point where each line is intersected by the negative diagonal. This procedure yielded d' values of 1·85, 1·66 and 1·58 for the Young, Middle and Old groups respectively. For statistical purposes, a value of d' was calculated for each subject by splitting his response matrix in such a way that his hit rate was approximately equal to his correct rejection rate. At this point, β is closest to 1·00 and the most stable estimate of d' is obtained. Mean values of d', calculated in this way were 1·99, 1·90 and 1·70 for the Young, Middle and Old groups respectively. These values were not significantly different from each other by the Mann-Whitney U test.

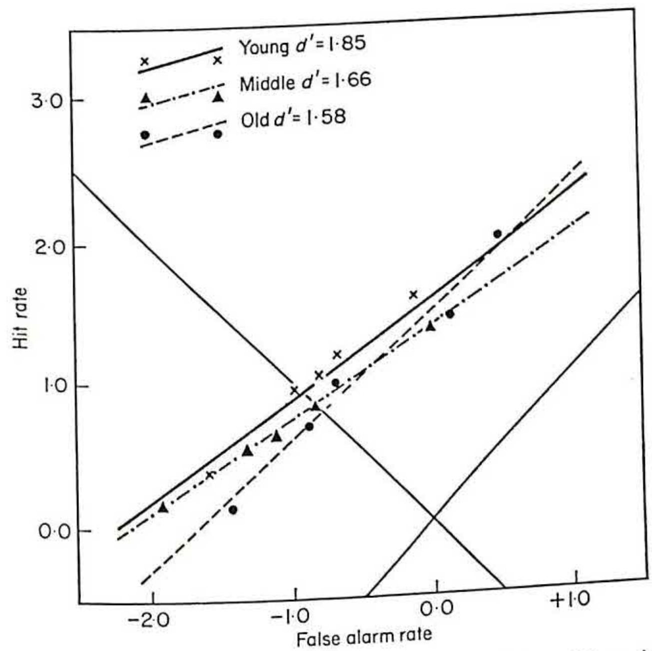

FIGURE 2. MOC curves for young, middle and old subjects (Experiment II).

Figure 2 shows that the data were quite well fitted by straight lines as they should be if SDT assumptions are valid. At the intersection of the lines with the negative diagonal (that is, where $\beta = 1{\cdot}00$), the curves were ordered Young, Middle, Old from top left to bottom right indicating that the highest d' value was for the Young group at that point. Mean d' values calculated from individual response matrices were also ordered Young, Middle, Old although the age differences were not statistically significant.

Thus, using the more sensitive techniques of signal detection analysis, t present experimental results confirmed Schonfield and Robertson's (1966) findi that there are essentially no age differences in word recognition performanc There is no evidence from the present experiment to support McNulty and Cair objections.

Discussion

Experiment I showed no significant age decrement in either acquisition or fo getting in a short-term recognition task. The second experiment demonstrated systematic, but non-significant, age decrement in secondary memory recognitio

With regard to the first (primary memory) task, it is worth noting that the mea age of the old subjects was over 76, yet there is no evidence for faster short-ter forgetting in that group. There is a systematic age difference in the acquisitio parameter although this effect was not statistically reliable. The results fro Experiment I thus demonstrate again that when a short-term memory task doe not demand division of attention or some manipulation of the stored material, ol subjects perform as well as a young group. A possible corollary of this conclusio is that older subjects are especially penalized by the divided attention situation an it is this fact, not a memory deficit as such, which led to the view that STM i impaired in the elderly. Evidence supporting the preceding statement comes fro a study reported by Schonfield (1969). He presented words dichotically and the gave subjects each word for recognition among three distractors. Schonfiel found an age decrement in recognition scores in both the first and the second half sets recognized. Thus apparently there *are* age differences in the short-term recognition of words when attention is divided at input.

The conclusion drawn from Experiment II is that recognition of words in secon dary memory is impaired only slightly. This result may be contrasted with the much larger age decrements in *recall* from secondary memory reported by Craik (1968) and Schonfield and Robertson (1966). The present results, together with the findings of Laurence (1967) and Schonfield and Robertson (1966), support the position that while there may be some age differences in registering verbal material in secondary memory, the major source of difficulty which older subjects encounter is located at retrieval.

This research was supported in part by a grant from the Medical Research Council. The author wishes to thank his colleagues at Birkbeck College for helpful comments on the manuscript.

References

BROADBENT, D. E. and HERON, A. (1962). Effects of a subsidiary task on performance involving immediate memory by younger and older men. *British Journal of Psychology*, **53**, 189–98.

BROMLEY, D. B. (1958). Some effects of age on short term learning and remembering. *Journal of Gerontology*, **13**, 398–406.

CRAIK, F. I. M. (1968). Two components in free recall. *Journal of Verbal Learning and Verbal Behavior*, **7**, 996–1004.

INGLIS, J. and CAIRD, W. K. (1963). Age differences in successive responses to simultaneous stimulation. *Canadian Journal of Psychology*, **17**, 98–105.
KEEVIL-ROGERS, P. and SCHNORE, M. M. (1969). Short-term memory as a function of age in persons of above average intelligence. *Journal of Gerontology*, **24**, 184–8.
KIRCHNER, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, **55**, 352–8.
KRIAUCIUNAS, R. (1968). The relationship of age and retention-interval activity in short-term memory. *Journal of Gerontology*, **23**, 169–72.
LAURENCE, M. W. (1967). Memory loss with age: a test of two strategies for its retardation. *Psychonomic Science*, **9**, 209–10.
MCNULTY, J. A. and CAIRD, W. (1966). Memory loss with age: retrieval or storage? *Psychological Reports*, **19**, 229–30.
MURDOCK, B. B. JR., (1966). The criterion problem in short-term memory. *Journal of Experimental Psychology*, **72**, 317–24.
SCHONFIELD, D. (1969). Recognition tests of dichotic listening and the age variable. Paper presented at XIX International Congress of Psychology, London.
SCHONFIELD, D. and ROBERTSON, B. (1966). Memory storage and aging. *Canadian Journal of Psychology*, **20**, 228–36.
TALLAND, G. A. (1965). Three estimates of the word span and their stability over the adult years. *Quarterly Journal of Experimental Psychology*, **17**, 301–7.
TALLAND, G. A. (1967). Age and the immediate memory span. *The Gerontologist*, **7**, 4–9.
TAUB, H. A. (1968). Age differences in memory as a function of rate of presentation, order of report, and stimulus recognition. *Journal of Gerontology*, **23**, 159–64.
TAUB, H. A. and WALKER, J. B. (1970). Short-term memory as a function of age and response interference. *Journal of Gerontology*, **25**, 177–83.
TUNE, G. S. (1966) Errors of commission as a function of age and temperament in a type of vigilance task. *Quarterly Journal of Experimental Psychology*, **18**, 358–61.
WELFORD, A. T. (1958). *Ageing and Human Skill*. London: Oxford University Press.
WICKELGREN, W. A. and NORMAN, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, **3**, 316–47.

Received 25 January 1971

Quarterly Journal of Experimental Psychology (1971) 23, 324–340

WAITING FOR THE STIMULUS SUFFIX: DECAY, DELAY, RHYTHM, AND READOUT IN IMMEDIATE MEMORY

ROBERT G. CROWDER
Yale University, Connecticut

The degree to which a redundant suffix impairs performance on digit lists depends on the delay between the last memory item and the suffix. A set of three experiments is offered, establishing the inadequacy of hypotheses based on simple decay of prelinguistic information, passage of time before recall, and rhythmic periodicities in attention. The delayed suffix effect is consistent with a fourth hypothesis whose emphasis is on access to precategorical auditory information through a "readout" process rather than on the sheer availability of such information.

Introduction

In a recent study on the suffix effect subjects were asked to recall series of vocally presented digits to which the word "zero" had been added following the last serial position. The zero suffix was a feature of every trial within a block of trials and did not have to be recalled. The subjects were told that the suffix was a cue as to when they should begin reproducing the memory series. The main variable in this earlier study (Crowder, 1969*a*) was the amount of time which elapsed between the last memory element and the suffix. The memory series were always presented at a rate of 2/sec, i.e. onset-to-onset time (OOT) of 0·5 sec. In one of the suffix conditions the suffix was presented in time with memory elements, at an OOT of 0·5 sec, whereas in the others it was delayed by OOT's of either 2·0, 5·0 or 10·0 sec. These four suffix conditions were compared with controls in which no suffix was presented and recall was immediate.

There were performance impairments in each of the four experimental conditions as compared with no-suffix controls; however, the pattern of results was quite different for the "immediate" (OOT of 0·5 sec) as opposed to the "delayed" (OOT of 2·0, 5·0 or 10·0 sec) suffix conditions. In the delayed conditions the impairment in recall was slight and not concentrated in any one portion of the serial position function. In the immediate condition there was a sharp increase in errors over the last three serial positions, virtually eliminating the recency effect.

Since elsewhere (serial positions 1–6) the delayed and immediate conditions did not differ, it was the case that performance varied directly with the passage of time before the recall cue, as in the phenomenon called *reminiscence*. Why forgetting should have been "backward" in this situation, with better recall coming after long as opposed to short retention intervals, is the question to which the present studies are directed. What is happening, in other words, while the subject is waiting for the suffix to occur?

The delay effect is consistent with a set of ideas advanced by Crowder and Morton (1969) about precategorical acoustic storage (PAS) (see also Morton, 1970). These authors maintained that when verbal stimuli reach the subject through his ears as opposed to only visually he has the advantage of a relatively longlasting prelinguistic store from which to extract information. This relative advantage of auditory over visual presentation is evident whether the two modalities are directly compared (Corballis, 1966; Routh, 1970) or whether silent visual presentation is compared with visual presentation accompanied by overt vocalization (Conrad and Hull, 1968; Crowder, 1970). Like the reminiscence effect obtained with a delayed suffix, the modality difference occurs only in the last few elements of the series. In fact the general rule is that an immediate suffix condition differs from its control condition in much the way a visual (control) condition differs from an auditory condition, viz. only slightly in the early positions but dramatically in the later positions. This circumstance led Crowder and Morton to suggest that both the suffix and the modality effects result from the PAS mechanism.

Their hypothesis was that immediately following auditory presentation the subject has an extra source of auditory information, coded only in gross physical features, from which he can extract information, or, as it were, "listen again" to the last few serial positions. This PAS information, however, is subject to degradation (either through overwriting or displacement) by subsequent receipt of similar auditory signals. Thus, when the suffix arrives it degrades the contents of PAS and removes some, apparently most (Crowder, 1969*b*), of the advantage of auditory presentation over visual. Serial position curves for the suffix vs. control comparisons and for the auditory vs. visual comparison look similar since each contains one function with benefits of PAS and each contains one without.

Just as the suffix is claimed to degrade PAS storage of the last few positions these earlier elements would logically have to be degrading information about *their* predecessors. The only elements relatively free from such retroactive effects would be those without any or many successors; thus, only they would be represented in PAS following the delivery of the last element. It is for this reason that the Crowder–Morton position is compatible with the fact that modality and suffix effects are restricted to the terminal list positions. As it happened, the restriction of PAS effects to late serial positions was compatible with a second hypothesis advanced in the earlier paper, and was thus somewhat overdetermined. The second limiting factor on PAS traces, besides displacement, was held to be temporal decay. The assumption was that PAS information decayed much more slowly than information from the comparable precategorical visual store (Averbach and Coriell, 1961), surviving on the order of a second or two. Given the timing of events in a memory span situation with a 2 digits/sec rate of presentation, by the time the last element has been presented PAS traces of the early serial positions would be too old to help in recall.

The delayed-suffix experiment was originally designed to test the decay hypothesis about PAS. The reasoning was that if the suffix effect is caused by elimination or degradation of PAS information then the suffix effect ought to vary in magnitude with the amount of information in PAS to be eliminated or degraded. If the suffix arrives immediately after stimulus presentation there should be a large

suffix effect since PAS is then presumed to be maximally full; however, after longer delays, during which PAS would be depleted through the decay process, the suffix should have little or no effect. This was the result obtained (Crowder, 1969*a*) and it matched nicely findings reported from studies in visual masking (Averbach and Coriell, 1961) where masking stimuli are more effective at short than at long inter-stimulus intervals.

The second outcome of the delayed-suffix experiment, the improvement in recall enjoyed by Subjects experiencing longer and longer postponements of the suffix, is also a feature of the visual masking paradigm. As there, this improvement must be attributed to a "non-selective readout" occurring between the sensory and memory stores. While this readout mechanism was not elaborated in the earlier paper, it now must be explicitly assumed that there is a transfer (either active or passive, see Routh, 1970) from PAS to some verbal short-term memory store beginning immediately after stimulus presentation. When no suffix occurs, the Subject completes the transfer and then offers his overt recall attempt; with a quickly delivered suffix, however, PAS undergoes degradation before this readout or rescue operation has been completed. This assumption constitutes recognition of the distinction between availability and accessibility (Tulving and Pearlstone, 1968) usually drawn mainly in experiments on long-term memory. Without drawing the distinction here, an apparent paradox arises in which better and better performance (with increasing suffix delays) is associated with less and less information (as the contents of PAS decay). While this is true of the raw availability of PAS traces, according to the decay hypothesis, it is also true that having traces available is little help if there is no "dead time" in which to identify them or read them out. With regard to the delayed suffix phenomenon, the real question is whether a decay mechanism is needed at all. For the reasons just cited (improved recall with increased suffix delays) it is not logically possible to hold a decay hypothesis without incorporating the principle of access through readout. It is, however, logically possible to maintain a readout hypothesis without an additional assumption about decay. It could, for example, be assumed that PAS traces last forever *unless* displaced by subsequent auditory events. The diminished suffix effect with long-delayed suffixes would then be explained by noting that the Subject had had adequate time to exhaust all the useful contents of PAS, although those contents were actually available until destroyed by the suffix. The third experiment in the present paper is directed at teasing apart implications of such a simple readout hypothesis from those of a readout-plus-decay hypothesis. Whichever account turns out to be correct, better data are needed on the form of the suffix delay function (Experiment I) as are tests of additional possible mechanisms different from both of those mentioned so far (Experiments II and III).

Experiment I

The original study showed the usual robust suffix effect at the "immediate" delay (OOT of 0·5 sec) and none whatever at the next longest interval of 2·0 sec. Experiment I was a replication intended to provide parametric information on what happens between these boundary conditions. The usual no-suffix control condi-

tion was compared to suffix conditions with (OOT) delays of 0·5, 0·6, 0·8, 1·0, 1·5 and 2·0 sec.

Method

Thirty-five Yale undergraduates listened to the same fixed list of 49 nine-digit series each presented at a rate of 2 digits/sec. These trials were broken into blocks of seven trials each, one condition per block. Three subgroups of subjects (n's = 14, 12, and 9) received the seven conditions in different orders according to a layout based on a randomized Youden square so as to minimize confounding of conditions across practice. Since individual stimuli were fixed to trial positions there was also some balancing of conditions among stimuli.

The conditions with delayed suffixes of 1·0, 1·5 and 2·0 sec were recorded directly on tape; timing of the suffix was facilitated by a metronome beating constantly and quietly in the background at a rate of 2/sec. The shorter delay conditions could not be estimated by a speaker at all accurately so all were recorded at a 1·0 sec delay and then adjusted by means of splicing out silent segments to achieve the desired OOT's for the suffixes. This could be done with excellent accuracy for the 0·8 and 0·6 sec OOT conditions. The shortest delay had originally been intended as a 0·4 sec OOT condition; however, since some of the digits in the memory series lasted longer than 0·4 sec, this goal could not have been accomplished other than at the expense of truncating memory stimuli (ninth position) in some cases. Of the 21 stimuli contributing to the shortest delay condition, 8 could be arranged for exactly the 0·4 sec OOT. Overall, the delays ranged from 0·4 to 0·58 sec with a mean of around 0·475 sec. This condition has simply been designated 0·5 sec for the present purposes.

The subjects were given general instructions as to their basic task at the onset of the experimental session. Also, before each block of trials they were given reminders relevant to the next set of tests (i.e. wait for suffix, etc.). The instructions specified delaying overt recall (always written) until the suffix occurred except in the single block of control trials where there was none. Since suffix delay varied across the conditions the suffix actually did serve the function of a recall cue. Each trial was preceeded by the word "ready" 1·0 sec before the series was presented. A new trial began every 30 sec, negligibly confounding the suffix-control comparison with recall time. Answer sheets providing nine spaces for each trial were provided for each block. The memory stimuli were permutations of the nine digits developed to minimize repetition of digits in the same position on adjacent trials and disallowing numerical sequences greater than two.

Results and discussion

The main data are given in Figure 1, which shows the six experimental conditions compared one at a time with the same (no-suffix) control data. While these curves are not quite as smooth as one would like, it is nonetheless apparent that the magnitude of the suffix effects exhibits a regular decline between OOT's of 0·5 and 2·0 sec. The reliability of this improvement in performance as a function of delayed recall is shown in Table I, which indicates separately for two performance measures the mean number of errors in each condition. Included also in Table I are the outcomes of statistical comparisons (two-tailed sign tests) among the various conditions—means not joined by the same underline were different at better than the 0·05 level of confidence.

Strictly speaking, the absolute number of errors in late serial positions is not the definitive measure of the suffix effect since the suffix might be raising error frequencies at all positions in the list. The *percentage* of all errors occurring on the

last position is a satisfactory measure of the shape of the function and is free from such objections. In Figure 2 are plotted error percentages as a function of suffix delay, with the control data represented by horizontal lines; for comparison purposes the comparable data of the previous experiment in this series (Crowder, 1969*a*) have been included. These data are perfectly consistent with those presented above in showing that the relative decrement in recency caused by the suffix dissipates with the passage of time since the last memory element.

The present data are thus in substantial accord with the earlier data and they show further that the function relating the suffix's effect to its delay is a gradual one. From Figure 2 one would not wish to claim that the function is asymptotic before OOT's of at least 2·0 sec and perhaps not even by then; however, the two-second figure derived from the earlier study (Crowder, 1969*a*) was not apparently grossly out of line.

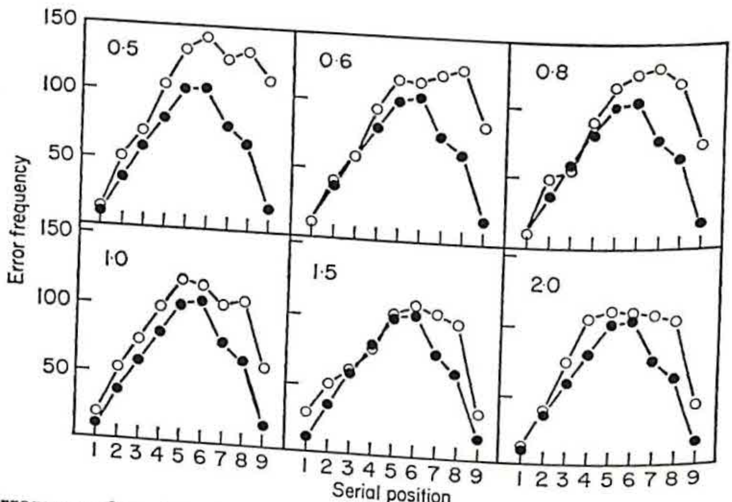

FIGURE 1. Errors as a function of serial position for each of six experimental (suffix) conditions (○—○) compared separately with the same control (no suffix) condition (●—●); the figure in the upper left portion of each graph is the OOT separating the suffix and ninth memory item.

TABLE I

Mean errors as a function of suffix delay

Measure	Condition 0·5	0·6	0·8	1·0	2·0	1·5	Control
Position 9	3·17	2·43	2·08	1·68	1·28	0·97	0·49
Positions 8 and 9	6·92	6·00	5·45	4·70	4·29	3·77	2·28

It is logically possible that the passage of time alone in the interval between stimulus presentation and recall is the variable producing data such as these. Since the seven conditions of Experiment I were all defined by different recall delays, there was a perfect confounding of suffix delay with recall delay. Although this *delay hypothesis* would have to be very complex (viz. to explain the sharp, position-specific decrement caused by the briefest 0·5-sec delay and then the gradual improvement in recall associated with the longer delays) the possibility was deemed worth checking in Experiment II.

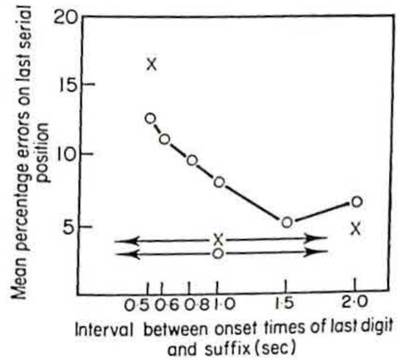

FIGURE 2. Mean percentage errors on the last position of suffix conditions as a function of OOT separating the suffix from the ninth position. Control (no suffix) data are represented by horizontal lines. ×, Crowder, 1969; ○—○, present experiment.

Experiment II

It has been shown that the suffix effect depends on a "physical match" between the acoustic channel over which the suffix comes and the acoustic channel over which the stimulus series comes (Crowder, 1969*b*; Crowder and Raeburn, 1970; Morton, 1970; Morton, Crowder, and Prussin, in press). Disparity on dimensions such as voice, pitch, laterality, etc., results in a reduced or attenuated suffix effect. Non-speech suffix noises give no suffix effect at all. This last property of the suffix effect (which is of course alleged to be a property of the PAS system) makes it easy to gather data on delayed recall in immediate memory where the suffix effect is not operating. The difference between conditions where a non-speech noise is used to mark the recall time and where speech suffix is used provides measure of suffix-delay effects uncontaminated by recall delay.

Method

Fifty Yale undergraduates listened to a list of 100 eight consonant series. The experimental design was a 5 × 5 factorial in which all comparisons were within subject. The first factor, and the one of interest here, was the recall condition: (a) no suffix; (b) a "zero" suffix at 0·5 sec delay (OOT from last memory element); (c) a "zero" suffix at 2·0 sec delay; (d) a buzzer suffix at 0·5 sec delay; and (e) a buzzer suffix at 2·0 sec delay. The second factor was concerned with the structure of the eight-letter stimulus items; a control condition with eight different consonant letters was compared with experimental conditions in which the second letter was repeated in position 4, 5, 6 or 7. Only data from the control condition will be considered here.* There were five blocks of 20 trials distinguished by recall conditions

*These data have been deposited with the American Documentation Institute.

in orders determined by a balanced latin square, thus balancing conditions against practice. Since a fixed list of stimuli was used, conditions were also balanced against individual memory series. The repetition conditions were mixed non-systematically with the control conditions such that four of each occurred within a block of 20 trials.

Stimulus items were spoken at a rate of 2 letters/sec with three trials given per minute. The buzzer used was a standard doorbell selected to be approximately as loud as the recorded voice and held down for approximately as long as it took to say "zero." In other details the procedure in Experiment II resembled that used in Experiment I.

Results and discussion

In Figure 3 the main data are given, with results from the buzzer conditions on the right and from the zero conditions on the left, both compared with the same no-suffix control data. When the word zero was used as the suffix the typical result was found. On the last serial position the 0·5 sec delayed suffix had a dramatic effect (as compared with control series) on errors ($P < 0{\cdot}01$ by two-tailed sign test), as did the suffix delayed by 2·0 sec ($P < 0{\cdot}01$); furthermore there was a significant difference at the last position between the two conditions involving the verbal suffix ($P = 0{\cdot}002$). On the other hand, there was no *selective* effect at all in

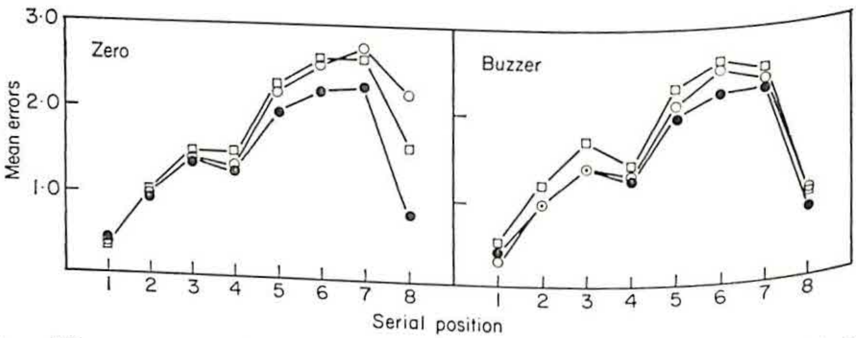

FIGURE 3. Mean errors as a function of serial position. In the left panel two suffix conditions with a speech sound delivered at different delays are compared with no-suffix control data. In the right panel, identical conditions are represented except that a buzzer is substituted for the speech suffix (the control data are the same in each panel). ●—●, Control; ○—○, 0·5 sec delay; □, 2-sec delay.

the conditions where a buzzer was substituted for the zero; neither the 0·5 sec nor the 2·0 sec buzzer conditions differed reliably from the controls on the last position nor did they differ from one another. (The earlier serial positions show non-selective effects of the buzzer which are not pertinent to the present discussions.) From these data we may conclude with confidence that the results of Experiment I and of Crowder (1969*a*) did not result merely from the passage of time, otherwise a comparable outcome should have been obtained in the buzzer conditions here.

A question of wider theoretical significance than elimination of the *delay hypothesis* is why the effects of speech and non-speech suffixes were so different. It would be too big a digression to consider this question in full; however, the two major theoretical alternatives will be briefly described since they apply in a somewhat more specific form to the effects of suffix-delay being discussed here.

The *structural* argument is that interferences at the PAS level occur because of the way the acoustic system is built. Inputs with different gross features get

handled in different parts of the system, and hence the buzzer does not act as a suffix because it gets entered in a different location to spoken digits. The *attentional* argument is that discriminations among gross features are used in determining what information is allowed *entry* into PAS. According to this view, anything allowed into PAS will act to degrade prior information, and it is assumed that the buzzer does not survive the very earliest stages of perceptual analysis and thus does not advance "as far into" PAS as the earlier memory stimuli are stored. These two hypothesized processes amount to a distinction between hearing and listening—a matching suffix has its effect simply because we hear it (structural) or rather because we are listening for it (attentional). These alternatives are similar to those which have been raised in the context of multichannel listening and selective attention (Moray, 1970; Norman, 1969; Treisman, 1969). The data needed to decide between the hypotheses, or possibly to reject both and choose a third, are not yet available, though early results with one technique (Morton, Crowder and Prussin, in press) seem to favor the attentional hypothesis.

At this point two explanations have been offered for the finding that a delayed suffix has a smaller effect than an immediate suffix. The *decay hypothesis* was originally advanced as part of the Crowder-Morton theory, but for logical reasons it was necessary to modify this hypothesis to include a *readout* process. This combined *decay-readout* hypothesis was challenged, in Experiment II, by the *delay hypothesis*, a contention that the time elapsing between the stimulus and permission to recall is the important variable; however, the results of Experiment II permit firm rejection of the delay notion. In addition, the structural and attentional arguments have been sketched, covering the general problem of trace interaction in PAS and the particular finding that a buzzer had no suffix effect (in Experiment II). The structural and attentional arguments will now be extended to cover the suffix-delay problem—an extension which was part of the rationale for Experiment III.

Experiment III

Experiment III was designed to test the possibility that the suffix effect occurs (with speech) to the extent the suffix is presented in time with the memory series. This notion, the *rhythm hypothesis,* will be recognized as a special case of the attentional argument. The idea is that the Subject is trained during presentation of the memory series that information of importance will be delivered to him every 0·5 sec, that he develops a periodic state of vigilance matching this tempo, and that such periodicity cannot be immediately turned off after the last memory element. Closure of some attentional system between periodic attention bursts would be adaptive in screening out distraction and in concentrating on subvocal rehearsal (Crowder, 1969*c*). Admission of an acoustically qualified suffix to PAS should, on this view, be substantially more likely the more that suffix occurs in rhythm with the memory series.

No data have been presented so far which could be used to deny that *all* suffix-delay effects arise solely from processes specified by the rhythm hypothesis since in all previous data the shortest suffix delay in addition to being the most potent

suffix condition was also the condition in which the suffix OOT matched that of the memory series.

Rhythmicity might thus be one dimension by which selective attention guards entry of information into PAS. The decay hypothesis is more properly considered as a version of the structural argument since decay is specified as an automatic and inherent property of the PAS system itself. The "critical" test of the attentional or rhythmic gating hypothesis is to refer to experimental conditions where OOT's during list presentation are slower than those separating the suffix from the last memory element. The briefly delayed arhythmic suffix should have a larger effect than the longer-delayed but rhythmic suffix, according to the decay hypothesis. A second requirement for the decay-plus-readout hypothesis (though not a differential prediction against the rhythm hypothesis) is that a suffix with a given fixed delay should have a larger effect the more rapid the presentation of the memory list. More precisely, it should have a larger effect on the next-to-last item when a rapid presentation rate had been used as opposed to a slow rate, Given that the suffix delay is fixed, by definition, the time elapsing between the very last item and the suffix would also be fixed. Thus, a fixed-delay ought to show an effect that extends *farther back* into the series if the list was presented rapidly than if it was presented slowly.

A further reason for conducting Experiment III was the availability of artificial (and automatically programmed) stimulus materials, assuring for the first time in this research that the sound of the memory materials was perfectly independent of rate of presentation and of suffix occurrence.

Method

The basic experimental design was a 3 × 3 × 2 factorial with three levels of presentation rate—OOT's of 0·25, 0·45 or 0·95 sec; three levels of suffix delay—OOT's (from the ninth element) of 0·25, 0·45 or 0·95 sec; and two types of suffix—a tone or the word "zero." Of the 56 college-age subjects (volunteers serving for pay) 20 received only control series (tone used as suffix) and 36 received only experimental series (zero used as suffix). All subjects served in all nine of the treatment combinations (three stimulus rates times three suffix delays) presented to subgroups of equal size according to a sequence dictated by the first four rows of a randomized latin square. Since exactly the same list of 63 stimulus items—seven per rate-delay block of trials—was used for every subject, the tone vs. zero comparison was perfectly independent of stimuli and the rate-delay comparisons were partly balanced both against stimuli and against stage of practice.

The special methodological feature of the present study was the use of a DDP computer (Honeywell Co.) to prepare the stimuli. A basic vocabulary for the study consisting of the nine digits, the words "zero" and "ready" and a 1000 Hz tone was prepared on the Haskins Laboratories Speech Synthesizer according to routines worked out by Mattingly (1968). Each of these utterances was then adjusted to have a duration of exactly 0·249 sec, comparable intensities, intonations, etc.* Assembly routines then permitted combining the elements of this vocabulary in different orders and in different temporal arrangements without in the least changing the acoustic properties of the stimuli from one sequence to another (see Aaronson, 1968 for a description of similar controls). The shortest interval possible between offset of one utterance and onset of the next was considerably less than

*The tone used in this study was actually an impoverished version of the synthetic "zero" used in the experimental conditions. Essentially, the second formant only was taken from the zero utterance and set at 1,000 Hz.

0·001 sec. Thus, the most rapid OOT was negligibly longer than 0·249 sec, say 0·25 sec. The other two values of OOT used in the present study were accomplished by inserting 0·200 or 0·700 sec between adjacent 0·250 sec utterances. Roughly these three rates correspond to 4/sec, 2/sec, and 1/sec respectively; the earlier data presented above suggest robust suffix effects ought to be obtained with acoustically qualified suffixes at these rates and delays.

The subjects were tested in groups ranging in size from one to five. Following general instructions they received some training in listening to synthetic speech. Between each of the nine blocks of seven trials a short break was called; however, since the recall interval (time between suffix offset and the ready signal for the next trial) was 20 sec, the whole experiment took less than 1 hr. In other respects the procedure here followed that used in the earlier studies on the series.

Results and discussion

(The results of a brief intelligibility test before the main experimental session indicated virtually perfect performance.) The means of errors at each serial position as a function of the 18 treatment combinations are given in Table II. Specific examination of the main hypotheses will be offered below; however, several overall analyses of variance were first prepared with the purpose of showing (a) that

TABLE II

Mean errors in Experiment III

Presentation rate (OOT)	Suffix delay (OOT)		Serial position 1	2	3	4	5	6	7	8	9
0·25	0·25	Control	0·40	1·40	2·55	2·80	4·15	4·65	4·95	4·05	0·65
		Suffix	0·42	1·62	2·58	3·08	4·03	4·72	5·33	5·20	4·47
0·25	0·45	C	0·15	1·80	2·20	3·35	4·35	5·00	4·65	3·75	0·30
		S	0·17	1·44	2·44	3·50	4·35	5·25	5·27	4·95	3·10
0·25	0·95	C	0·20	1·75	3·30	3·65	4·60	5·30	5·25	3·20	0·40
		S	0·44	1·42	2·58	3·56	4·41	5·06	5·34	4·50	2·56
0·45	0·25	C	0·10	1·55	2·35	2·75	3·45	4·05	4·15	3·20	0·55
		S	0·42	1·34	1·75	2·30	3·35	4·10	4·16	4·63	3·43
0·45	0·45	C	0·15	1·45	2·55	2·65	2·90	3·90	4·40	3·40	0·40
		S	0·39	1·16	2·38	2·91	3·82	4·02	4·71	4·46	3·88
0·45	0·95	C	0·55	1·55	2·25	2·75	3·55	3·55	4·15	3·50	0·80
		S	0·80	1·80	2·72	3·35	3·96	4·41	4·64	4·43	2·85
0·95	0·25	C	0·25	1·20	1·90	2·15	2·45	3·25	3·70	2·55	0·75
		S	1·03	1·88	2·78	3·02	3·71	4·07	4·32	4·57	3·08
0·95	0·45	C	0·35	1·45	2·25	2·70	2·75	3·45	4·10	3·30	0·55
		S	0·78	1·50	2·02	2·60	3·22	3·68	4·16	4·39	3·18
0·95	0·95	C	0·45	1·50	2·05	2·60	3·20	3·40	3·75	3·05	0·60
		S	0·83	1·79	2·30	2·73	3·46	3·66	3·78	4·02	2·88

the data were generally sensitive to the experimental treatments, and (b) that the outcome was consistent with previous research insofar as it was a replication and refinement of earlier procedures.

One expectation from the preceeding study is that whereas rate of presentation might make a difference in the control experiment (i.e. for the 20 subjects who consistently received a tone as the suffix) delay of the suffix should not. The analyses of variance shown in Table III reveal this to have been the case. The

TABLE III

Analyses of variance for Experiment III

Source	Controls (tone) Positions 1–7 *df*	MS	*F*	8–9 MS	*F*	Suffix (zero) Positions 1–7 *df*	MS	*F*	8–9 MS	*F*
Presentation rate	2/38	549·04	12·14*	4·41	1·31	2/70	325·69	5·92*	21·91	4·32*
Suffix delay	2/38	51·51	1·49	0·11	<1·0	2/70	39·08	1·24	60·11	9·80*
Interaction	4/70	18·76	<1·0	5·01	1·60	4/140	107·07	5·09*	10·98	1·92

*$P<0·01$.

main effect of presentation rate was statistically significant for serial positions 1–7 in the control data but not the main effect of suffix delay or the interaction. (Inspection of the data indicated that the one way of collapsing serial position without serious loss of information was the 1–7, 8–9 scheme used here.) Also, if the suffix effect has the same properties with synthetic speech as with natural speech we should expect (a) that suffix delay should affect performance only when a speech sound is used as the suffix and, (b) that this effect should occur only in late serial positions.

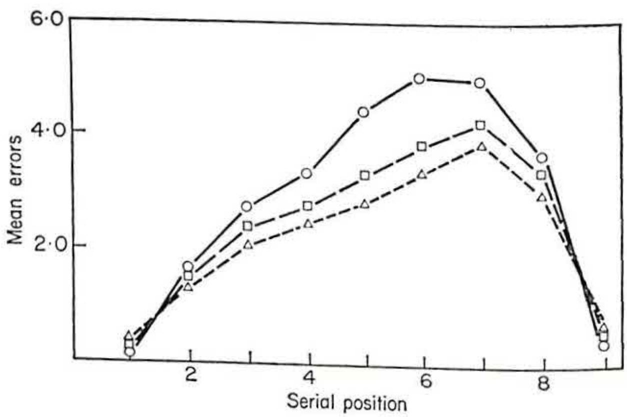

FIGURE 4. Mean errors as a function of serial position for the control (tonal suffix) conditions. These data have been collapsed over tone-delay. Presentation rate (OOT in sec): ○—○, fast (0·250); □ - - □, medium (0·450); △ - - △, slow (0·950).

These expectations are amply confirmed in the data of Table II and the analyses of Table III. Suffix delay had a statistically significant effect only in the experimental (speech suffix) condition and only for the last two serial positions. Thus the data from the present study are in broad agreement with the previous findings regarding suffix effects and are also of sufficient reliability to justify conducting individual comparisons among conditions of special interest.

The control data were next examined to see whether collapsing across suffix-delay conditions would be permissible, as the low F ratios in Table III for this factor suggest. Two-tailed sign tests among delay conditions within a given presentation-rate condition gave no statistically significant results at all for positons 1–7 and only one statistically significant result among the data for positions 8 and 9—at the fastest presentation rate more errors were made following a 0·950 sec delay of the tone (OOT) than following an immediate (OOT of 0·250 sec) tone ($P < 0{\cdot}01$). Since this unpredicted difference was not reflected in the overall analyses of Table III and since a theoretical interpretation has not, at the time of writing, suggested itself the decision has been made to ignore it and accordingly the control conditions have been collapsed across suffix-delay conditions in Figures 4 and 5. The control data are shown in Figure 4 with presentation rate the parameter.

In agreement with earlier work where stimuli have been appropriately controlled against confounding of intelligibility and presentation rate (Aaronson, 1968) performance is shown to be better with slower rates than with faster rates and the difference is found in the interior serial positions.

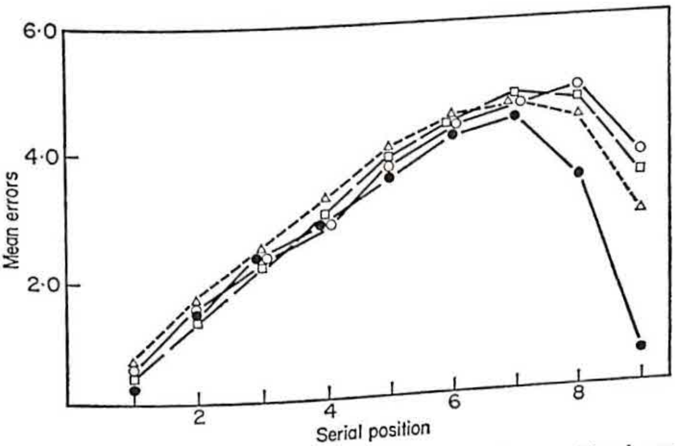

FIGURE 5. Mean errors as a function of serial position. The combined control (tonal suffix) data (●—●) are compared against conditions differing in suffix delay (OOT in sec): ○—○, 0·250; □- - □, 0·450; △- -△, 0·950

The data on delayed suffixes are given in Figure 5, where all the control data have been collapsed into a single function as have experimental (speech suffix) conditions differing only in presentation rate. These combined data show that overall delaying the suffix had the same effect here as in previous work; performance on early serial positions was left unchanged but performance on the last two positions was inversely related to the delay after which the suffix was delivered. The data of Figure 5 do not, however, answer the major theoretical question posed in the present study: are suffix delay effects caused by the delay itself or by a confounding of delay with the rhythm in which the series is presented? Evidence on this point is presented in Figure 6, which breaks down the difference among the experimental groups several ways. In each of the three panels three functions have been drawn relating errors to suffix delay with presentation rate the parameter.

Panel (a) gives the mean error frequencies for the experimental group alone on position 9; panel (b) gives these frequencies as proportions of the total number of errors made in a condition; and panel (c) shows z-scores based on Mann–Whitney tests comparing the control and experimental groups within each rate–delay combination. Although the latter are generally used for inferential purposes they are necessarily monotonic in the "size" of an experimental effect. While considerable erudition might be brought to bear on which of these performance measures is the "best" one, the fact is that the issue cannot be resolved at the present state of quantification in theory except that panel (b) probably has the edge since PAS theory insists on changes in the *shape* of the serial position curve rather than absolute differences.

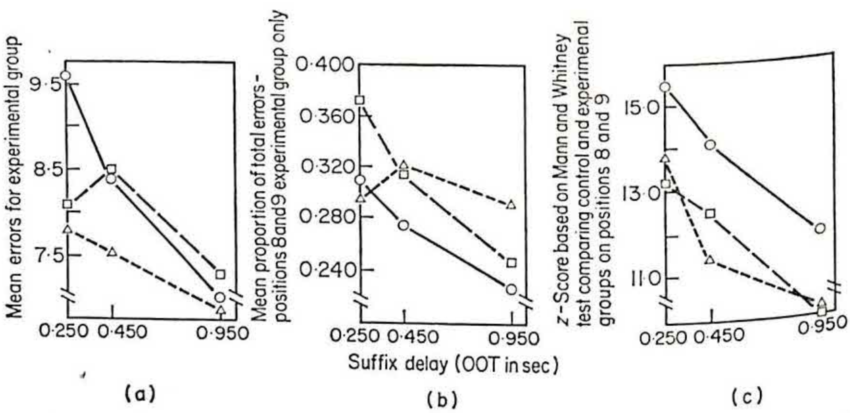

FIGURE 6. Effects of suffix delay upon errors in late serial positions plotted separately for the three presentation rates and given in three different types of performance measure. Presentation rate (OOT in sec): ○—○, 0·250; □- -□, 0·450; △- -△, 0·950.

Fortunately the three versions of Figure 6 are not too much in disageement. All show a reduced suffix effect with increased delays; however, there is some discrepancy between panels (a) and (b) regarding whether the function is invariably monotonic. The two conditions which are most crucial are those in which the presentation rate was either moderate (OOT of 0·450) or slow (OOT of 0·950) since in the fast rate the most rapid suffix condition was also the one which delivered a suffix in rhythm with the memory series. According to the decay-plus-readout hypothesis all three curves should be monotonic while according to the rhythm hypothesis the peak suffix effect at the moderate presentation rate ought to be with a suffix delay of 0·450 sec and the peak suffix effect with the slow rate ought to be at the longest delay (OOT of 0·950) of the suffix. The data of panel (a) for intermediate presentation-rate condition provide the only comfort in Figure 6 for the rhythm hypothesis. Aside from this non-monotonicity and the inconclusive data on the slowest presentation rate in panel (b), the results are unanimous in support of the decay hypothesis. The analyses of variance in Table III showed furthermore no interaction between presentation rate and suffix delay, as the rhythm hypothesis would require. A comparable analysis of variance for the data of panel (b) showed a non-significant main effect of presentation rate, F (2, 70) = 2·54, and a non-significant interaction, F (4, 140) = 2·38, but in agreement with the other

analyses a strong main effect of suffix delay, $F(2, 70) = 10{\cdot}91$, $P < 0{\cdot}01$. (The significant but unpredicted main effect of presentation rate for experimental subjects on positions eight and nine in Table III does not survive the proportional transformation.)

Experiment III has so far thus appeared to support the decay-readout hypothesis more than the rhythm hypothesis. Crowder and Morton (1969, Experiment I) showed that the size of the suffix effect (in running memory span) was directly related to the speed of presentation. This they interpreted as being consistent with the decay prediction that the faster the presentation rate the more information ought to be present in PAS for the suffix to interfere with. However, in the Crowder-Morton study the presentation rate was perfectly confounded with suffix delay, i.e. the suffix was always in rhythm with the memory series, therefore the larger effect at fast rates could be a product of the shorter delay separating the memory series from the suffix. The present study permits an unambiguous look at this proposition.

The reasoning behind an interest in comparing the effects of constant-delay suffixes following varied presentation rates is as follows: either or both of two factors could be limiting duration of information in PAS–temporal decay or limited capacity in terms of space. If space limitations alone are operating, the contents of PAS may be estimated by knowing how many items have been delivered through the ears (with additional stipulations about the "channels" over which this information has come). The other possibility, the original Crowder–Morton position that both decay and space limit the contents of PAS, is that given an equal numerical count going into PAS, the mechanism will be fuller if information has been rapidly received than if it has been received slowly. Thus, when a suffix is presented at a fixed OOT, say 0·500 sec, the decay hypothesis would require a larger effect of the suffix if the last three or four elements had been read in at a rate of 4/sec than at a rate of 1/sec. In a limiting case, where items were delivered at a *weekly* rate we would expect no effect of the suffix on items except perhaps the last. In fact, since the delay between the suffix and the last memory item is being held constant, the predictions being developed here must be applied only to the penultimate and perhaps antepenultimate items.

Figure 7 shows data permitting inference on whether a fixed-delay suffix has effects extending earlier into the series following fast presentation rates that follow slow presentation rates. Since it is not permissible to collapse control conditions across presentation rates (see Figure 4), difference scores between corresponding experimental and control conditions have been plotted. There is certainly no indication in the data of Figure 7 that a constant-delayed suffix is having either a larger effect after fast presentation rates than after slow presentation rates or an effect that extends farther back into the series. Crowder and Morton (1969) were apparently mistaken in supposing that fast presentation rates exposed more information to PAS disruption than slow rates.

Returning to the original question concerning why a delayed suffix has a smaller effect than an immediate one, the justification for a decay hypothesis has been considerably weakened. If one wished to claim that the differences between ninth-position performance in Figure 5 occurring as a result of delaying the suffix

either 0·250 or 0·950 sec comes from decay of information in PAS, then why in Figure 7 does not *eight-position* performance show similar variation when equally wide variations in suffix delay are being compared? There are two consequences of abandoning the decay principle: (a) attribution of all limitations on the capacity of PAS to space, and (b) attribution of the delayed-suffix effect entirely to the readout process. In other words, after an arbitrarily long (silent) period, say 6 months, after stimulus input, the suffix will still have just as drastic an effect on the *availability* of PAS information as after a fraction of a second; however, this delayed suffix effect on PAS availability will not be observable since *access* to the information has long since been accomplished. The choice anticipated earlier between a decay-readout and a readout-only hypothesis has been settled in favour of the latter.

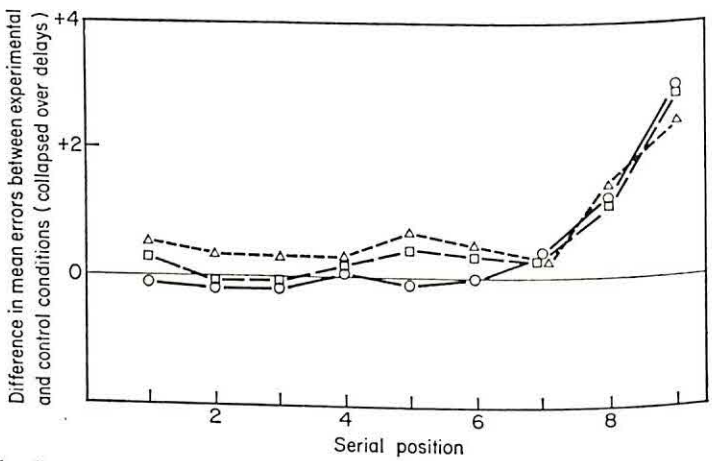

FIGURE 7. Absolute differences in mean errors between experimental and control conditions as a function of serial position with presentation rate the parameter. Conditions differing only in suffix delay have been combined. Presentation rate (OOT in sec): ○—○, 0·250; □- - □, 0·450; △- -△, 0·950.

There is one logical obstacle to this interpretation and it is a major one. *If* it is assumed that the readout process occurs not only at the end of stimulus presentation but also between items during presentation, then predictions from the decay-plus-readout hypothesis and the readout-only hypothesis become identical. Taking again the limiting case where the memory series is given at a weekly rate, it is seen that if the subject rescues the PAS information from each item between items, there should be a suffix effect only on the last position (providing there is a brief suffix delay). On the other hand, a rapid presentation rate would mean a great deal of rescue work remains to be done following the last item and therefore a suffix should interfere seriously with performance. This reasoning seems difficult to challenge, and it is hard to imagine finding an effect in the eighth serial position, such as in Figure 7, if the items had been received as slowly as one every five sec or so. However, it is possible that in fact the subjects always postpone readout of the last two or three items until presentation is complete when rates within the range used here are employed. Such a strategy would actually be designed to capitalize on PAS, or, generally, to capitalize on transient, ready-access memory for recent

events as opposed to more permanent but more highly-encoded memory for earlier positions. Some support for the suggestion that the subjects allow loading of the last few positions before readout is given by the data of Figure 4, where, in the control conditions, presentation rate had no effect on the last two positions. Also, studies where the subject has been allowed to select his own study time for each serial position (Belmont & Butterfield, 1969) have shown rapid input of the last few positions, presumably directed at capitalizing on "primary" or prelinguistic memory.

General Discussion

Experiments II and III permitted reasonably straightforward rejection of two hypotheses which could logically have been used to account for the delayed suffix effect: (a) a complex relation between performance, serial position, and recall delay; and (b) an hypothesis that the subjects can ignore suffixes increasingly well as they increasingly violate the rhythmicity used in presenting the memory items. The alternatives left after eliminating these possibilities included the original decay hypothesis modified to include a readout process, or, alternatively, the readout hypothesis alone. In the experiments reported here there have been no results requiring a decay principle beyond the readout principle so therefore the tentative conclusion has been reached that PAS is not time-limited, at least not for the materials and time values used in the present type of research.

One conjecture on how a directed readout system would work is that between input of the memory list and recall the subject goes through a subvocal "dress rehearsal" at a rapid rate. This silent dress rehearsal can be seen as the last stage of a cumulating rehearsal strategy the subject uses during the earlier portions of list input (Crowder, 1969*c*). The time at which PAS information could be used would be just as the subject reaches the last few elements in his dress rehearsal. As he articulates these elements subvocally he can simultaneously consult the contents of PAS to determine whether a match exists, and if not, recategorize the discrepant element. Whether PAS information is autonomously available during this period and even whether it is as well being autonomously fed into later parts of the system does not matter as long as there can be directed readout (a more or less voluntary comparisons process) just as the subject reaches the last part of the list in his silent rehearsal loop.

Now since the dress rehearsal process is claimed to be occuring at the articulatory level it must be happening at the sluggish rate of processing associated with speech (Landauer, 1962). This in turn means that between the time the subject hears the last memory element of a nine-digit series something like a second or two will elapse before he reaches the point in his dress rehearsal when PAS would be concluded. Naturally, if a suffix occurs during this period to spoil the PAS traces, the main advantage of auditory presentation will have been lost. However, if there is time to make the crosscheck between his rehearsal loop and PAS before arrival of the suffix then the later spoilage of PAS information is largely irrelevant since access has already taken place. It is of particular interest that the time estimate made above—that it will take between one and two seconds of subvocal articulation to reach the

last one or two memory items—corresponds conveniently to the time intervals over which the suffix exerts its effect on performance.

This research was supported by NSF Grants GB 4066 and GB 15157. Portions of the data from experiments I and II were presented at the Psychonomic Society meeting, St. Louis, 1969. The assistance of Vicki Raeburn, Henry L. Roediger III, and James Antognini in collecting and tabulating data for this paper is acknowledged with appreciation.
I wish to thank Ruth S. Day for her generous assistance in preparing the stimulus materials; her help and the hospitality of Haskins Laboratories were indispensible for Experiment III.

References

AARONSON, D. (1968). Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology*, **76**, 129–40.
AVERBACH, E. and CORIELL, A. S. (1961). Short-term memory in vision. *Bell Systems Technical Journal*, **40**, 309–28.
BELMONT, J. and BUTTERFIELD, E. C. (1969). The relations of short-term memory to development and intelligence. In LIPSITT, L. C. and REESE, H. W. (Eds.), *Advances in Child Development and Behavior*, Vol. 4, pp. 71–72. New York: Academic Press.
CONRAD, R. and HULL, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science*, **10**, 135–6.
CORBALLIS, M. C. (1966). Rehearsal and decay in immediate recall of visually and aurally presented items. *Canadian Journal of Psychology*, **20**, 43–51.
CROWDER, R. G. (1969*a*). Improved recall for digits with delayed recall cues. *Journal of Experimental Psychology*, **82**, 258–62.
CROWDER, R. G. (1969*b*). An auditory subsystem in human memory. Paper presented at the 136th meeting of the American Association for the Advancement of Science, Boston.
CROWDER, R. G. (1969*c*). Behavioral strategies in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, **8**, 524–8.
CROWDER, R. G. (1970). The role of one's own voice in immediate memory. *Cognitive Psychology*, **1**, 157–78.
CROWDER, R. G. and MORTON, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics*, **5**, 365–73.
CROWDER, R. G. and RAEBURN, V. P. (1970). The stimulus suffix effect with reversed speech. *Journal of Verbal Learning and Verbal Behavior*, **9**, 342–5.
KAHNEMAN, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, **70**, 404–25.
LANDAUER, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, **15**, 646.
MATTINGLY, I. (1968). Synthesis by rule of general American English. Suppl. to *Status report on speech research*. Haskins Laboratories.
MORAY, N. (1970). *Listening and Attention*. London: Penguin.
MORTON, J. (1970). A functional model for memory. In NORMAN, D. A. (Ed.), *Models for human memory*. New York: Academic Press.
MORTON, J., CROWDER, R. G. and PRUSSIN, H. A. Experiments with the stimulus suffix effect. *Journal of Experimental Psychology*, in press.
NORMAN, D. A. (1969). *Memory and Attention*. New York: Wiley.
ROUTH, D. A. (1970). "Trace strength", modality, and the serial position curve in immediate memory. *Psychonomic Science*, **18**, 355–7.
TREISMAN, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, **76**, 282–99.
TULVING, E. and PEARLSTONE, A. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, **5**, 381–91.

Received 5 May 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 341–345

GENERALIZED AND LATERALIZED EFFECTS OF CONCURRENT VERBALIZATION ON A UNIMANUAL SKILL

MARCEL KINSBOURNE AND JAY COOK

Division of Pediatric Neurology, Duke University Medical Center, Durham, North Carolina

After practice, subjects balanced a dowel rod on the right and on the left index finger while speaking and while remaining silent. As compared to control, the verbal condition yielded shorter balancing times for the right hand, but longer ones for the left. A speculative model postulates enhancement of the practised skill by virtue of the distraction effect of the concurrent activity. This is counteracted on the right by interference with right-sided motor control by the left cerebral hemisphere due to verbal activity programmed by the same hemisphere.

Introduction

The human observer, working at full capacity, will lose efficiency if required simultaneously to perform a second task (Fitts and Posner, 1967; Taylor, Lindsay and Forbes, 1967; Welford, 1968).

The type of performance usually tested in "Time Sharing" experiments that show this effect is one that incorporates tasks involving constant monitoring control, such as tracking an unpredictably moving target (Trumbo, Noble and Swink, 1967). Practised tasks, for which a motor program can be worked out during acquisition trials, and subsequently run off at will (Fitts and Posner, 1967) are not necessarily interrupted by concurrent activity—women can talk while knitting—(Bahrick and Shelly, 1958). May such a task, even under appropriate circumstances, be facilitated by concurrent activity? The possibility has been suggested (Bliss, 1892; Boder, 1935) but never demonstrated experimentally (Keele, 1968) except in the special case of concurrently heightened muscle tension (Bills, 1927) which increased speed of certain reactions, perhaps the result of greater motor preparedness (Courts, 1942).

The length of time for which a dowel rod may be balanced on one finger increases with practice, the preferred hand maintaining an advantage. A pattern of coordinated wrist and arm movements is acquired against a background of occasional readjustments of body position. This unimanual task affords the opportunity of comparing the effects of concurrent activity separately on performance with the right and with the left hand.

The left cerebral hemisphere, which also controls movements of the right hand, subserves language function in right-handed subjects. Though the corpus callosum unites the cerebral hemispheres, right-sided control of the left hand is

relatively well removed anatomically from the left-sided locus of language behaviour. The use of verbal intercurrent activity affords the opportunity of comparing the effects of a lateralized secondary task on ipsilaterally controlled (right-handed) and contralaterally controlled (left-handed) main performance. Will "hemispheric sharing" between concurrent activities generate mutual interference between them, with performance decrement proportionately greater than to any that occurs between activities located in different hemispheres?

We examined the effect of concurrent verbalization on balancing a dowel rod on the right and on the left index finger.

Materials and Methods

Subjects

Twenty normal subjects were used, aged 19–20 years. All had shown firm right-hand preference on an unpublished item check list with questions referring to habitual hand preference in a number of activities, each rated on a five point scale. Two were rejected from the data analysis because they failed to meet a pre-set criterion of a mean balancing time in excess of 5 s for the 20 experimental trials.

TABLE I

Median balancing times (sec)

	LN	LV	RN	RV
Subject 1	88	114	106	55
2	107	143	112	64
3	96	191	85	85
4	77	115	85	115
5	83	125	89	125
6	12	14	36	20
7	12	10	11	9
8	12	15	14	13
9	10	6	12	8
10	7	9	18	14
11	4	6	11	7
12	5	10	7	7
13	13	16	30	16
14	19	15	31	17
15	40	17	45	43
16	30	40	98	35
17	9	9	13	13
18	29	32	37	61
Mean of medians	36·3	49·3	46·7	39·3

Procedure

Subjects practised balancing a wooden dowel rod (1·3 cm by 46 cm long, weighing 105 g), while in the standing position for 15 min, sharing time equally between the two hands. They were asked to try to keep the rod balanced as long as possible, making whole body movements if necessary. They balanced alternately on each hand for 1 min at a time with

1 min intermissions. They then proceeded to the experimental trials, which included a non-verbal and a verbal condition. The non-verbal condition was as practised. The verbal condition differed only in that it incorporated the additional requirement that prior to each act of balancing, the subject listen to a sentence, then repeat it aloud continually (without correction) while balancing the dowel rod.

Twenty trials were given, counterbalanced between hands (right and left) and between the two conditions (verbal and non-verbal) as follows: RN, LN, RV, LV, repeated five times for nine subjects, and LV, RV, LN, RN, for the other nine. For each trial the length of time for which the subject was able to keep the dowel rod balanced on one finger was recorded. A trial was concluded if balancing time reached 250 s.

Results

There was significant variance between conditions ($\chi r^2 = 13{\cdot}25$, $P < 0{\cdot}01$) in a Friedman analysis of variance.

Wilcoxon tests revealed: the right hand performed better under the non-verbal condition ($T = 13{\cdot}5$, $P < 0{\cdot}01$). The hands did not differ under the verbal condition ($T = 55{\cdot}5$ NS). The right hand performed less well under the verbal as compared to the non-verbal condition ($T = 33$, $P < 0{\cdot}05$). The left hand performed better under the verbal as compared to the non-verbal condition ($T = 32$, $P < 0{\cdot}05$).

Discussion

Contrary to the usual outcome of time sharing experiments, concurrent verbalization actually enhanced dowel balancing on the left index finger. Thus, time sharing does not necessarily impair a practised skill. What might the mechanism of enhancement be?

Weinstein and MacKenzie (1966) obtained improved manual performance in the presence of white noise. This improvement was plausibly attributed to the resultant arousal when the main task in monotonous, long drawn out and makes very low demands on capacity (Antrobus, 1964). Does the requirement to verbalize have an arousing effect? Dowel balancing is done over short periods of time and is an exacting task, and presumably quite "arousing" itself. In any case, arousal cannot account for a differential effect between the hands. The difficulty of the main task also makes it implausible to resort to the "challenge effect" as the source of improvement. This is an enhancement of performance due to increased motivation when an easy task is made more difficult. Again, this does not predict the decrement in right hand performance. The enhancement effect of muscle tension on speed of response (Courts, 1942) does not generalize to the present findings, as muscle tension actually has an adverse effect on postural stability (Seashore and Koch, 1938). Perhaps we should seek an explanation in a closer analysis of the attentional requirements of a practised task.

While a skill is being acquired, attention in the form of observing response is fully focused upon its components, distraction is poorly tolerated and all available sources of feedback are used in building up the appropriate motor programmes. Once this is achieved, the attentional requirements slacken (Posner, 1969), and possibly the reprocessing of feedback information now interferes with the sequence

of events by inducing intermittency. It may help not to focus attention on a practised action pattern, for fear of disrupting it (as the pianist falters who watches his fingers). Perhaps verbalization enhanced performance because it was just sufficient to distract attention from the main task, but not demanding enough to pre-empt all the available capacity. The result would then illustrate the following generalizations:

Execution of a practised skill benefits from relatively undemanding concurrent activity but suffers from concurrent activity of a more exacting kind (Fig. 1).

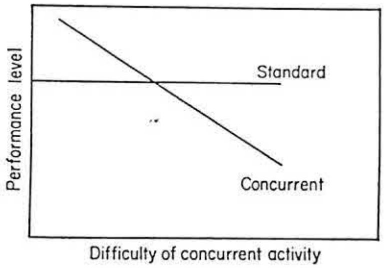

FIGURE 1. Effect of concurrent activity as a function of its difficulty.

In the course of practice, the effect of an undemanding concurrent task changes from hindrance to help (Fig. 2).

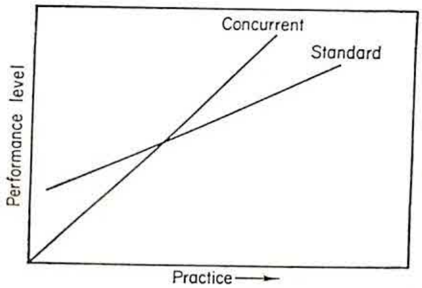

FIGURE 2. Effect of concurrent activity as a function of practice.

It remains to be considered why the right hand did not share the enhanced performance of the left. Had the performance decrement been bilateral, one might have argued for a general effect, such as the switch from the practised silent balancing situation to the novel condition of balancing while speaking. However,

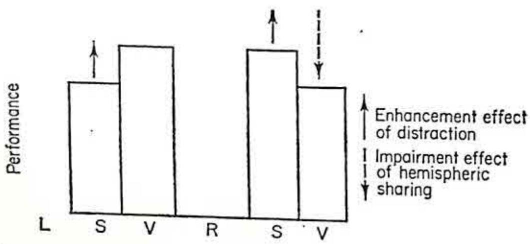

FIGURE 3. Two-factor model to account for results.

the effect was lateralized. A further interacting variable, with asymmetric impact, has to be identified. It could be that mutual interference between main and secondary task in the left cerebral hemisphere ("hemisphere sharing") counteracts the enhancement effect for the right hand, and sets up an overall decrement in performance on that side. The postulated interactions are depicted in Figure 3.

We are grateful to Dr Richard Pew of the Human Performance Center, Ann Arbor, Michigan, for helpful discussion. This study was supported by UPHS—National Institute of General Medical Sciences—5T01-GM-01238.

References

ANTROBUS, J. S. (1964). Visual signal detection as a function of sequential variability of simultaneous speech. *Journal of Experimental Psychology*, **68**, 603.

BAHRICK, H. P. and SHELLY, C. (1958). Time sharing as an index of automatization. *Journal of Experimental Psychology*, **56**, 288–93.

BILLS, A. G. (1927). The influence of muscular tension on the efficiency of mental work. *American Journal of Psychology*, **38**, 227–51.

BLISS, C. E. (1892–93). Investigations in reaction time and attention. Studies from the Yale Psychological Laboratory, **1**, 1–55.

BODER, D. P. (1935). The influence of concomitant activity and fatigue upon certain forms of reciprocal hand movement and its fundamental components. *Comparative Psychology Monographs*, **11**, 4.

COURTS, F. A. (1942). Relations between muscular tension and performance. *Psychological Bulletin*, **39**, 347–62.

FITTS, P. M. and POSNER, M. I. (1967). *Human Performance*. Belmont, Calif.: Brooks-Cole.

KEELE, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, **70**, 387–403.

POSNER, M. I. (1969). Reduced attention and the performance of "automated" movements. *Journal of Motor Behavior*, **1**, 4, 245–58.

SEASHORE, H. G. and KOCH, G. (1938). Postural steadiness under conditions of muscular tension and fatigue. *Psychology Record*, **2**, 319–32.

TAYLOR, M. M., LINDSAY, P. H. and FORBES, S. M. (1967). Quantification of shared visual capacity processing in auditory and visual discriminations. *Acta Psychologia*, **27**, 223–9.

TRUMBO, D., NOBLE, M. and SWINK, J. (1967). Secondary task interference in the performance of tracking tasks. *Journal of Experimental Psychology*, **73**, 232–40.

WEINSTEIN, A. and MACKENZIE, R. S. (1966). Manual performance and arousal. *Perceptual Motor Skills*, **22**, 489.

WELFORD, A. T. (1968). *Fundamental of Skill*. New York: Barnes and Noble, Inc.

Received 14 *May* 1971

BOOK REVIEWS

POSTMAN, L. AND KEPPEL, G. (Eds.). *Norms of Word Association.* New York and London: Academic Press. 1970. Pp. vi + 467. £12·00.

This book comprises a collection of nine sets of association norms, most of which were not previously available, except as technical reports. The first four sets include the Russell-Jenkins (1952) norms, together with responses to the Russell-Jenkins stimuli of British, Australian, German and French students, and French workmen. Rosenzweig discusses some of the cultural differences, but apart from a stronger tendency for English-language speakers to give the dominant associate these differences do not appear to be either very large, or obviously meaningful.

The Russell-Jenkins norms were replaced by the more extensive Palermo-Jenkins norms in 1964. Those have already been published and are hence not included. However, Keppel and Strand give norms for associations to the Palermo-Jenkins associations, which allows an estimate of which associations are bidirectional, and also enables three-step associations to be produced. Postman reports the California norms for two-syllable nouns from four ranges of Thorndike-Lorge frequency. Marshall and Cofer report associations to the dominant responses from 21 of the Connecticut category norms, and Cramer lists the associations to 100 homographs, words with identical spelling but two meanings (e.g. pupil: student, or part of the eye). These are useful in giving an indication of the relative probability of the two meanings. Ervin-Tripp tries to bridge the gap between word associations and connected speech by following her free association test with a request to construct sentences containing the stimulus word, and to produce words that could be substituted for the stimulus word. Each set of norms is accompanied by a description of the procedure for collection and usually by a discussion of the results.

Word association norms clearly reflect strong verbal habits, which have a powerful effect in many verbal learning tasks, and it is easy to see why a simple response measure strongly suggesting the Hullian concept of habit strength should have seemed so promising to Russell and Jenkins in 1952. It is much more difficult in 1971 to share Jenkins' enthusiasm and belief that "the free association technique may well be instrumental in unlocking the secrets of verbal behaviour and, indeed, perhaps of language itself" (p. 7). It is hard to believe that a measure that lumps together such diverse responses as synonyms, antonyms, superordinates and subordinates can ever be a sharp enough tool to attack the thorny problems of semantics. In the case of connected language, the value of free association techniques seems even less; as Ervin-Tripp points out, "associative strength in fact is a very poor device for prediction of sequences in discourse" (p. 394). However, it is sometimes useful to take advantage of the verbal habits reflected by association norms, regardless of one's theoretical predelictions, and for that reason it is useful to have published norms. At £12 a copy, this book is not likely to head the list of best sellers, but it is probably sufficiently useful to be bought by most laboratories working in the area of verbal learning and memory.

A. D. BADDELEY

WOLSTENHOLME, G. E. W. AND KNIGHT, J. (Eds.). *Taste and Smell in Vertebrates.* London: J. & A. Churchill. 1970. Pp. x + 402. £4·00.

Technical problems of stimulus presentation and theoretical problems of stimulus classification are particularly severe for taste and smell. Yet that is not a good enough excuse for physiologists' and experimental psychologists' relative neglect of these modalities, one of which may be the most important of all for many vertebrates. Only now are straightforward ultrastructural and biochemical studies of receptors being produced,

although these have been technically possible for over a decade. A few workers have undertaken electrophysiological or degeneration investigations over the years, and there have been a few useful lesion studies. It is to be hoped that this symposium will persuade some up-and-coming bright and physiologically educated psychologists that in these areas there are now rich pickings to be had for real science, and that it is not just a question of following your nose as a matter of taste.

Although there are larger recent symposium volumes covering the field, this book provides a conveniently brief impression of recent work. It contains the papers and discussions from a 1969 Ciba Foundation symposium attended, as usual, by a small but fair selection of current authorities from Europe and North America.

The six papers on gustation deal with the tongue and with the mind but not with the brain. Pfaffmann and Zotterman update their laboratories' reports on glossopharyngeal and chorda tympani impulse frequencies in animals and man, in relation to behavioural, psychophysical and experimential responses. The concept of four taste primaries receives critical discussion. Receptor interactions are covered very briefly. Recent advances in the understanding of taste bud ultrastructure are described in detail, and the sensory cell tentatively identified. There are some extraordinary scanning electron microscope pictures of taste buds.

Cerebral mechanisms are better represented for olfaction, but then the peripheral mechanisms are even less clear than for gustation and so an understanding of how the sensory modality operates is just as distant. There are two papers on electron microscopy of the olfactory bulb and a review by Adey on "higher olfactory centres", including EEG and unit recordings against a background of experimental anatomy. At the receptor level we are given more electron microscopy, an autoradiographic study of cell proliferation in the olfactory epithelium and a paper on the electro-olfactogram. To complete the picture there are five papers playing the familiar game of constructing receptor interaction theories out of structure–activity relationships (Amoore, Beets, Davies, Døving, Wright). Also Ashton and Eayrs present their old study of the dog as mine-detector, presumably as an entertaining indication of how defence money should be spent. It seems that if your dog keeps on retrieving an old bone from the rose bed or the dustbin, the thing to do is bury it (the bone) under a large lawn and roll the turf.

D. A. Booth

Tobias, J. V. (Ed.). *Foundations of Modern Auditory Theory*, Vol. I. New York and London: Academic Press. 1970. Pp. xv + 466. £10·50.

Since Licklider's chapter in S. S. Stevens' *Handbook of Experimental Psychology*, no comparable review of psychoacoustics (or preferably, "the psychology of hearing") has appeared. Too much work has been under way to permit any researcher the time, the wide grasp and the relatively stable state of the art requisite for anything approaching a text. Tobias has produced the only practical solution by obtaining chapters of from 30 to 50 pages from eleven specialists. So great has been the need for such a standard source on hearing that it hardly matters that some of the chapters are already dated.

Small's review of periodicity pitch phenomena is particularly welcome, since no previous article has adequately summarized the psychological evidence on harmonic and anharmonic sounds as well as the physiological evidence. Clear distinctions are drawn between possible mechanisms for periodicity pitch, envelope detection, fine structure detection, etc. While a lot of the detail we need is not yet available, the impressive battery of evidence defies the prejudiced to continue dismissing periodicity pitch as an artifact.

Many writers, the Signal Detection Theorists (SDT) included, have at some time been unclear as to whether Signal Detection Theory is about attention and decision-making or about the auditory system. In an otherwise excellent review, Jeffress shows that SDT is a good tool in auditory psychophysiology as elsewhere, but fails to warn readers against taking its concepts and its parameter estimates too literally. No musical party-goer should be without W. Dixon Ward's review of consonance, vibrato and special pitch phenomena. Just intonation, equal temparament and Pythagorean tuning need never again be confused,

and the hi-fi pundit can be told that 5% harmonic distortion is undetectable (by whom?). This lively writer should have been let loose on musical aesthetics as well; with the existing *lacuna* in the area of perception of structured auditory material other than speech such a chapter would certainly be justified in a book of this kind.

Auditory science has become much less speculative, even in the last 20 years. No major shifts of theoretical affiliation are revealed in this book. The idea of "neural sharpening" has fallen slightly from favour. Chiefly von Bekesy's idea, this invoked a process of lateral inhibition similar to that operating in Mach Bands in vision to explain the fine frequency discrimination possible despite the extremely broad-band frequency analysis offered by the basilar membrane travelling-wave envelope. Von Bekesy's own emphasis has shifted to the temporal basis of inhibitory phenomena apparently operating in the spatial (i.e. frequency) domain. F. Blair Simmons points out that one early suggestion of sharpening via progressively narrower neural tuning curves at ascending levels of the auditory system has not been confirmed; the best stimulating frequency in the threshold sense and the preferred frequency or frequencies for neural period synchrony may be different functional attributes of the same fibre. Although psychological demonstrations have recently been offered that lateral inhibitory processes do play a part in the sharpening of frequency discrimination, physiological work by de Boer (1967) shows finely tuned fibres in cat auditory nerve when no lateral inhibition is possible, namely, when the stimulus is white noise. We are back with Licklider and Huggins' principle, that the nervous system will do things as many different ways as it possibly can. Tonndorf's chapter on cochlear mechanics gives the most plausible of the basically physical explanations of fine frequency discrimination yet advanced; in the region of the maximum displacement for a given frequency there is a switch-over from mainly radial to mainly longitudinal shearing forces between basilar and tectorial membranes. Shear is the ultimate stimulus to the hair-cells and hence likely to be important. If effectively "radial" and "longitudinal" inhibition operate in the two-dimensional grid at or behind the hair cells, then frequency-sharpening can be explained.

None of the chapters in the book are weak. Von Bekesy's exposition and experimental methods remain cavalier but he has earned this privilege; his place-analysis bias is countered by a time-analysis bias on the part of Nordmark. Elliott and Fraser on fatigue and adaptation, Teas on cochlear electrophysiology, and Schuknecht on auditory lesions are more empirically oriented. Scharf's comprehensive review of the "critical band" literature may tend to be relatively undervalued because earlier reviews exist. Tobias' one-paragraph introductions tell us colloquially but tactfully what to expect from each contributor.

Hearing is rarely taught in psychology courses, beyond the fragmentary, uninteresting and outdated story available in single chapters of books on the senses. With the publication of this book the two classical reasons have evaporated; no longer is recent literature inaccessible, and no longer is the psychology outweighed by physical and physiological groundwork. "Foundations" is more for the teacher than the pupil; all the more reason to teach rather than refer. But hearing is a large and difficult subject and there is still room for a modern popular treatment, without which, I fear, the teaching will not occur.

MARK P. HAGGARD

Reference

DE BOER, E. (1967). Correlation studies applied to the frequency resolution of the cochlea. *Journal of Auditory Research*, **7**, 209–17.

HERMELIN, B. AND O'CONNOR, N. *Psychological Experiments with Autistic Children*. Oxford: Pergamon Press. 1970. Pp. 142. £3·00.

This is a model book. It is an account of a patient and extraordinarily versatile attempt to tackle a difficult and often unrewarding problem. Autistic children are a recent phenomenon, but their existence, once recognized, excited a very wide interest. This is natural because there is always the hope that the handicap will be curable, if enough is found out about it. The hope is an admirable one, but has unfortunately tempted most of the

researchers in this field to oversimplify and to explain the condition of these unhappy children in terms of some current bandwagon. Some, for example, say autistic children are autistic because they are under-loved, others because they are over-aroused.

Hermelin and O'Connor have successfully avoided the temptation to squeeze autism into any preconceived and simplistic frameworks such as these, and, indeed, one important feature of their book is their demonstration that attempts to do so are wrong. The book describes a series of experiments on perception, learning, language, concept formation, EEG patterns and social behaviour in autistic children. Their approach from the start was an open one, and they have produced some varied and important results. For example, autistic children have difficulties with language, a general deficit which has always been recognized. Yet the authors demonstrate that their ability to attend to and store verbal material presented in a rote fashion is quite as good as equivalent normal children. On the other hand, autistic children produce startlingly inferior performances when the material is presented in a connected and meaningful way. Similarly, it has been suggested that autistic children are over-aroused, and, therefore, withdraw from external stimulation. Yet this book demonstrates, through a series of telemetric EEG studies, that the differences between autistic and other children are a great deal more intricate than this. Autistic children are more aroused than others by the onset of sound, but there is no equivalent effect for the onset of light. On the other hand, they habituate to the presence of light stimuli more rapidly than other groups. It is, of course, essential in studies such as these that the precise nature of the stimuli should be systematically varied, and it is to the great credit of the authors to have instituted these variations.

The book should be read by a wide range of psychologists. It is a convincing demonstration of the value of experimental techniques, sensitively and systematically applied.

P. E. Bryant

Black, P. (Ed.). *Physiological Correlates of Emotion.* New York and London: Academic Press. 1970. Pp. xvi + 309. £6·30.

This book is yet another collection of chapters by different authors which is based on a conference, in this case one held in Baltimore in 1968. Seven papers presented at the conference serve as the nucleus for the volume which contains, in addition, six other chapters specially invited to broaden the total scope. The result is a comprehensive and up-to-date coverage with a fair uniformity of presentation. We are spared those sometimes entertaining, but usually irritating, "verbatim" reports of discussion following the presentation of papers. The book is divided into five sections, the first of which is the Introduction and contains a single chapter by Bindra. The next section is entitled, "Genetic and Developmental Correlates", and comprises a paper by Bruell, and, inevitably, one from the Harlows. Well known names also appear in Part 3 which is entitled, "Neurochemical and Endocrine Correlates"; these are Kety, Grossman and Brady. The next part, on the Neurophysiological Correlates of Emotion, has chapters by MacLean, Lindsley, and Delgado; while the final part, Psychophysiological Correlates, contains contributions from the Laceys, Valins, Hammond and Arnold.

It is supposed that a book such as this might serve at least two functions, firstly as a general textbook for undergraduate teaching, and secondly, as a more advanced text for postgraduates, academics and research workers to bring them up-to-date with the present state of the subject.

There has for a long time been a dearth of modern, reasonably comprehensive, tests which could be used for undergraduate teaching, and this volume goes a long way to filling that need. Bindra's excellent opening chapter provides just the right sort of setting which can help to orientate the undergraduate reader to an adequately broad non-dogmatic view of the subject. His advocacy that emotion should not be treated as something uniquely different from other aspects of behaviour, although reiterating the stand taken by others over the years, is nevertheless still welcome. The continuance of work in the area loosely described as "emotion", however, recognizes that most people understand what is meant

by the term and no doubt it will continue to be used as a convenient coverall necessitating reminders like this latest one of Bindra's that the area is not uniquely defined.

The next chapter by Bruell, on the heritability of emotional behaviour, is somewhat specialized and the coverage is not wide. It does, however, contain a useful section starting off with the provocative statement "it is a common misconception that, by definition, genetically determined traits are also heritable". The following chapter by the Harlows, while a necessary requirement of any text that aims at comprehensiveness, has a ring of familiarity. It is as always, interesting, clear and well illustrated.

Kety's chapter is short, but in the space available he presents a clear picture of a fairly wide coverage of neurochemical aspects. His contribution is perhaps best seen as a condensed, and slightly more up-to-date version of his well-known paper with Schildkraut. In contrast, Grossman's chapter on the modification of emotional behaviour by intracranial administration of chemicals appears to lack incisiveness by perhaps attempting too encyclopaedic a coverage. Undergraduates would not enjoy this, but those working in the field would probably fare better. Brady's chapter is clear, useful, but predictable, and has the usual illustrations.

MacLean admits that he strays slightly from his assignment to explore work on the psychoses in relation to the limbic brain. His deviation is worthwhile, and while illustrating points pertinent to the general topic of the book he provides a coverage which is topical in the light of "current" developments. Lindsley is concerned with the brain area with which his name is usually connected and he draws together much of the newer work and relates it to older well-established lines of country. He appropriately finds it useful to discuss such topics as selective attention and vigilance in this context. Delgado's contribution is concerned with a short presentation of his well-known work on radio stimulation. The spectacular nature of these studies tends to overshadow their potential value in freeing experimental subjects from the confines of the normal laboratory.

The contribution of the Laceys might not be thought of as strictly within the terms of reference of the book, but is clearly within the ambit if emotion is reviewed from a wider standpoint; their point of view on the relations between autonomic, cortical and perceptual events is well-known and while presenting some new data from their laboratory, it is a pity that they do not take into account the work of other experimenters which closely impinges on theirs. Valin's chapter is perhaps one of the most central to the theoretical interpretation of emotional behaviour in man, insofar as it is concerned with the labelling of bodily changes. The experiments which he presents are intriguing and perhaps because of this deserve to be fully replicated in a way that those of Schacter, from whom the work is derived, have not been. Hammond's chapter on conditioned emotional states starts with a consideration of the Mowrer, and Rescorla and Solomon position and extends study from there to work on "relief and disappointment". This is welcome in an area where coverage in reviews has been too long concerned with fear and anger. The final chapter is by Magda Arnold and presents an up-dated version of the theoretical position with which she is associated.

In summary, a worthwhile book, a curate's egg with far more good than bad. It will serve both as an undergraduate textbook, albeit probably of most value to the finalist, and as a stimulus to the flagging academic and retarded researcher.

P. H. VENABLES

BAIRD, J. C. *Psychophysical Analysis of Visual Space.* Oxford: Pergamon Press. 1970. Pp. viii + 321. £4·50.

Physical dimensions are usually specified with reference to one-, two- or three-dimensional systems of co-ordinates. For present purposes, these may be understood to extend linearly throughout space and all to have the same metric properties. The congruence of linear measurements, irrespective of the directions in which they are made, is perhaps the most important characteristic feature of physical space. Upon it depend the theoretical principles of geometry and trigonometry and their application to all forms of linear, area and volumetric measurement. A human observer may be regarded as situated at the origin

of his own individual coordinate system also. He reacts to and interacts with stimuli appearing within that system in accordance with information encoded by his senses, particularly his eyes. Much of man's everyday behavioural commerce with his surrounding environment depends upon relatively acute detection of the presence of visual stimuli and reasonably accurate observation of their locations, shapes and sizes. The degree of precision to which man's repertoire of sensory, perceptual and instrumental skills can be developed testifies strongly to the extent of apparent agreement between physical and perceptual dimensions. Thus, the spatial senses and the brain mechanisms associated with them usually succeed in providing a remarkably accurate account of the form, structure and content of the environment. With so-called "naïve-realism", therefore, the layman supposes his "sense-data" to be both valid and reliable.

In contrast with physical space, however, the metric properties of visual space are "anisotropic"; that is to say, spatial dimensions appear to differ according to their orientations and the directions in which they are made. Accordingly, some of the most significant chapters in the history of experimental psychology have been written in the attempt to establish relationships between physical dimensions and subjective experience of them. The book under review is the latest in a long line of such attempts. It begins what is described as, "An analysis of visual space" by referring to the "perceptual constancies" and the "invariance hypothesis" (p. 1). Issue is taken over conventional treatments of these topics, however, since, "As with the discussion of constancy, no mention is likely to be made about experimental data which are consistent with the invariance hypothesis". Therefore, "this book was written to provide an alternative viewpoint on the problems of visual space". This alternative viewpoint "is that of psychophysics, applied in an objective way to several problems of visual space", namely, "the problem of size and distance" (p. 2). The classical psychophysical tradition demands first that the stimulus is specified definitively. The assumption adopted in analysing the several types of spatial judgement referred to in subsequent chapters of this volume is, "that the critical variables in visual perception are proximal, involving angular relations at the eye" since, "In the present work it is assumed that measurements should concern the relations among retinal images". Accordingly, considerable pains are taken throughout to convert the results of studies, made largely in Sweden and North America, from metric into angular units. Unfortunately, although it purports to "concern the relations among retinal images", by restricting consideration of proximal representation of distal stimulus-patterns to the "visual angles subtended by metric stimuli at the frontal corneal surface" (p. 3), this analysis neglects consideration of the spatial cues available within the eye itself. Most previous attempts to develop a systematic account of spatial psychophysics have been guilty of similar neglect. Since overlooking actual proximal dimensions has generated much confusion over the principles of space perception it is timely to reconsider the nature of relationships between the sizes of stimulus-objects, the angles subtended by their boundary-rays as they converge towards the eye and the dimensions of images projected by them upon the retinal mosaic.

When Descartes analysed the cues available for perceiving the sizes of distant stimuli he made two basic, but fallacious, assumptions about image dimensions. The first implies that the retina is essentially flat and thus analogous to a photographic film. The second suggests that images developed upon the retinal surface alter directly in magnitude, and linearly, so with those of a variable stimulus situated at a fixed distance. It also follows from the second assumption that images vary inversely in size for a given stimulus at varying distances. These assumptions are adopted in all the basic texts on space perception and they have been repeated so frequently as to have become almost tantamount to "a priori" principles. They seem equally profound when translated into visual angle terms also. Thus, when the author of this volume states that, "For frontal size at a constant distance, linear metric-size is directly proportional to subtended visual angle; whereas, metric distance of a frontal target is inversely proportional to its visual angle" (p. 243) he appears to be adopting the traditional assumptions relating stimulus and image dimensions. But there are several confusions here. First, the retinal surfaces of the eye is not flat. Moreover, after a moment's reflection, possibly with the aid of a simple sketch, readers will

by the term and no doubt it will continue to be used as a convenient coverall necessitating reminders like this latest one of Bindra's that the area is not uniquely defined.

The next chapter by Bruell, on the heritability of emotional behaviour, is somewhat specialized and the coverage is not wide. It does, however, contain a useful section starting off with the provocative statement "it is a common misconception that, by definition, genetically determined traits are also heritable". The following chapter by the Harlows, while a necessary requirement of any text that aims at comprehensiveness, has a ring of familiarity. It is as always, interesting, clear and well illustrated.

Kety's chapter is short, but in the space available he presents a clear picture of a fairly wide coverage of neurochemical aspects. His contribution is perhaps best seen as a condensed, and slightly more up-to-date version of his well-known paper with Schildkraut. In contrast, Grossman's chapter on the modification of emotional behaviour by intracranial administration of chemicals appears to lack incisiveness by perhaps attempting too encyclopaedic a coverage. Undergraduates would not enjoy this, but those working in the field would probably fare better. Brady's chapter is clear, useful, but predictable, and has the usual illustrations.

MacLean admits that he strays slightly from his assignment to explore work on the psychoses in relation to the limbic brain. His deviation is worthwhile, and while illustrating points pertinent to the general topic of the book he provides a coverage which is topical in the light of "current" developments. Lindsley is concerned with the brain area with which his name is usually connected and he draws together much of the newer work and relates it to older well-established lines of country. He appropriately finds it useful to discuss such topics as selective attention and vigilance in this context. Delgado's contribution is concerned with a short presentation of his well-known work on radio stimulation. The spectacular nature of these studies tends to overshadow their potential value in freeing experimental subjects from the confines of the normal laboratory.

The contribution of the Laceys might not be thought of as strictly within the terms of reference of the book, but is clearly within the ambit if emotion is reviewed from a wider standpoint; their point of view on the relations between autonomic, cortical and perceptual events is well-known and while presenting some new data from their laboratory, it is a pity that they do not take into account the work of other experimenters which closely impinges on theirs. Valin's chapter is perhaps one of the most central to the theoretical interpretation of emotional behaviour in man, insofar as it is concerned with the labelling of bodily changes. The experiments which he presents are intriguing and perhaps because of this deserve to be fully replicated in a way that those of Schacter, from whom the work is derived, have not been. Hammond's chapter on conditioned emotional states starts with a consideration of the Mowrer, and Rescorla and Solomon position and extends study from there to work on "relief and disappointment". This is welcome in an area where coverage in reviews has been too long concerned with fear and anger. The final chapter is by Magda Arnold and presents an up-dated version of the theoretical position with which she is associated.

In summary, a worthwhile book, a curate's egg with far more good than bad. It will serve both as an undergraduate textbook, albeit probably of most value to the finalist, and as a stimulus to the flagging academic and retarded researcher.

P. H. Venables

Baird, J. C. *Psychophysical Analysis of Visual Space.* Oxford: Pergamon Press. 1970. Pp. viii + 321. £4·50.

Physical dimensions are usually specified with reference to one-, two- or three-dimensional systems of co-ordinates. For present purposes, these may be understood to extend linearly throughout space and all to have the same metric properties. The congruence of linear measurements, irrespective of the directions in which they are made, is perhaps the most important characteristic feature of physical space. Upon it depend the theoretical principles of geometry and trigonometry and their application to all forms of linear, area and volumetric measurement. A human observer may be regarded as situated at the origin

of his own individual coordinate system also. He reacts to and interacts with stimuli appearing within that system in accordance with information encoded by his senses, particularly his eyes. Much of man's everyday behavioural commerce with his surrounding environment depends upon relatively acute detection of the presence of visual stimuli and reasonably accurate observation of their locations, shapes and sizes. The degree of precision to which man's repertoire of sensory, perceptual and instrumental skills can be developed testifies strongly to the extent of apparent agreement between physical and perceptual dimensions. Thus, the spatial senses and the brain mechanisms associated with them usually succeed in providing a remarkably accurate account of the form, structure and content of the environment. With so-called "naïve-realism", therefore, the layman supposes his "sense-data" to be both valid and reliable.

In contrast with physical space, however, the metric properties of visual space are "anisotropic"; that is to say, spatial dimensions appear to differ according to their orientations and the directions in which they are made. Accordingly, some of the most significant chapters in the history of experimental psychology have been written in the attempt to establish relationships between physical dimensions and subjective experience of them. The book under review is the latest in a long line of such attempts. It begins what is described as, "An analysis of visual space" by referring to the "perceptual constancies" and the "invariance hypothesis" (p. 1). Issue is taken over conventional treatments of these topics, however, since, "As with the discussion of constancy, no mention is likely to be made about experimental data which are consistent with the invariance hypothesis". Therefore, "this book was written to provide an alternative viewpoint on the problems of visual space". This alternative viewpoint "is that of psychophysics, applied in an objective way to several problems of visual space", namely, "the problem of size and distance" (p. 2). The classical psychophysical tradition demands first that the stimulus is specified definitively. The assumption adopted in analysing the several types of spatial judgement referred to in subsequent chapters of this volume is, "that the critical variables in visual perception are proximal, involving angular relations at the eye" since, "In the present work it is assumed that measurements should concern the relations among retinal images". Accordingly, considerable pains are taken throughout to convert the results of studies, made largely in Sweden and North America, from metric into angular units. Unfortunately, although it purports to "concern the relations among retinal images", by restricting consideration of proximal representation of distal stimulus-patterns to the "visual angles subtended by metric stimuli at the frontal corneal surface" (p. 3), this analysis neglects consideration of the spatial cues available within the eye itself. Most previous attempts to develop a systematic account of spatial psychophysics have been guilty of similar neglect. Since overlooking actual proximal dimensions has generated much confusion over the principles of space perception it is timely to reconsider the nature of relationships between the sizes of stimulus-objects, the angles subtended by their boundary-rays as they converge towards the eye and the dimensions of images projected by them upon the retinal mosaic.

When Descartes analysed the cues available for perceiving the sizes of distant stimuli he made two basic, but fallacious, assumptions about image dimensions. The first implies that the retina is essentially flat and thus analogous to a photographic film. The second suggests that images developed upon the retinal surface alter directly in magnitude, and linearly, so with those of a variable stimulus situated at a fixed distance. It also follows from the second assumption that images vary inversely in size for a given stimulus at varying distances. These assumptions are adopted in all the basic texts on space perception and they have been repeated so frequently as to have become almost tantamount to "a priori" principles. They seem equally profound when translated into visual angle terms also. Thus, when the author of this volume states that, "For frontal size at a constant distance, linear metric-size is directly proportional to subtended visual angle; whereas, metric distance of a frontal target is inversely proportional to its visual angle" (p. 243) he appears to be adopting the traditional assumptions relating stimulus and image dimensions. But there are several confusions here. First, the retinal surfaces of the eye is not flat. Moreover, after a moment's reflection, possibly with the aid of a simple sketch, readers will

realize that even if it were image dimensions cannot be enlarged linearly with equal incremental increases in stimulus size. Secondly, regardless of the shape of the retina, or of the refracting properties of the optical mechanisms provided that they remain constant, changes in visual angles are necessarily reflected in correspondingly proportional changes in images. This follows from the photo-optical properties of composite lens systems of the kind embodied within the human eye. Nevertheless, it does not follow that, "at a constant distance linear metric size is directly proportional to subtended visual angle". This statement is only true for stimuli appearing upon the inner surface of a sphere within which the observer's (single) eye is situated always at the centre.

For many, the need for a thoroughgoing "psychophysical analysis of visual space" is one of the most outstanding priorities for experimental psychology. It must start with a sophisticated understanding of the relationship between distal and proximal dimensions and proceed by establishing how spatial judgements vary correspondingly. This may be expected to provide the framework for a new theory of space perception which, hopefully, will not conclude by stating that, "the basis of an observer's distinctions among the absolute metric locations, shapes and sizes of targets must depend upon factors other than those considered in the present analysis" (p. 298). Regrettably, such a theory remains to be developed.

GERALD H. FISHER

FLORES D'ARCAIS, G. B. AND LEVELT, W J. M. (Eds.). *Advances in Psycholinguistics.* Amsterdam and London: North Holland. 1970. Pp. x + 454. £8·40.

Advances in Psycholinguistics contains a selection of papers given at a conference held at Bressanone under the auspices of the University of Padua in July 1969. The book exhibits both the virtues and drawbacks of published proceedings, particularly of "workshop" gatherings of specialists engaged in active research in a rapidly advancing field. While the enthusiasm engendered by exchange of ideas, methodologies and experimental results comes across, there is the difficulty not just of the inevitable time-lag in publication of experimental data but the problem of keeping pace with changing conceptions about the status of psycholinguistic research. The editors' aim is to concentrate on reports of empirical research. Consequently, there is a slightly uneasy lack of contact between the discussion of basic assumptions in the few theoretical contributions and the majority of experiments reported which tend to take these same assumptions for granted.

Among the experimental papers, of particular interest are those describing work carried out in European countries which might not otherwise easily be available to English speaking psychologists. Examples are Levelt's use of scaling methods to measure syntactic structure on the basis of subjects' judgements of relatedness between words in Dutch sentences; Parisi and Antinucci's semantic analysis of spacial locatives in Italian; Flores d'Arcais' experiments with Italian comparatives; Legrenzi's report on the effects of expressing a reasoning rule using different constructions in Italian; and De Boysson-Bardies' investigation of French children's memory for explicit and implicit negatives. The importance for language research of data from different languages should be obvious, a point explicitly made in Slobin's paper on linguistic universals based on cross-linguistic similarities in children's acquisition of languages as varied as German, Russian, Finnish, Samoan and Luo. These and other detailed reports of research make this a valuable source-book for psycholinguists interested in specific problems such as the perception of ambiguity, the ability to evaluate different types of sentences and the stages by which children achieve mastery of certain constructions.

However, when one comes to consider the overall impact of the book, the editors' attempt to impose some kind of theoretical framework highlights the difficulty of categorizing these experiments according to the suggested criterion of whether linguistic structure is the independent or dependent variable. This raises two issues, the first of which is discussed in a characteristically trenchant paper by Bever on the interactions between linguistic and perceptual abilities. His contention is that people's intuitions about the unacceptability of utterances are often based not on deviance from grammatical rules but on perceptual

complexity. He further goes on to claim that grammatical rules themselves may have developed in a particular form in order to avoid perceptual confusability. While his arguments are persuasive, they leave unsolved the basic issue of the extent to which linguistic usage is constrained by non-linguistic factors such as perceptual and memory limitations. If this is the case, the question is whether the limitations should be written into a grammar which aims to give an account of linguistic competence as displayed in usage of a language. One can understand the reluctance of linguists to allow for the interference of non-linguistic variables that cause temporary fluctuations in language performance. But can they afford to ignore systematic features of human processing abilities, particularly in view of Chomsky's claim in *Language and Mind* that linguistics is a branch of cognitive psychology because it provides information about the organization of human mental processes?

The other basic issue, reflected by the categorization of experiments into sections on grammar, lexical structure and meaning, and cognition and language, is only touched on incidentally in most of the papers. At bottom this is the still unresolved problem of the relation between language and thought, which in psycholinguistic terms becomes the interaction between grammar and meaning. When a psychologist studies subjects' recall of comparative sentences or children's ability to comprehend them, is he investigating the logical or semantic manipulation of concepts such as greater or smaller or the effect of expressing these concepts in different linguistic forms? Johnson-Laird raises this issue in his paper on quantifiers such as *all*, *some* and *many*, questioning the extent to which the results of linguistic analysis can be translated into logical propositions. It is certainly difficult, if not impossible, to draw the line between experiments on cognition, meaning and grammar, since the effects of syntactic factors are usually measured in terms of subjects' comprehension of sentences while reasoning processes are tested using various syntactic constructions. The central problem that runs through the book is the basic psycholinguistic dilemma: how to isolate the role of syntactic rules when their only function is to communicate semantic content.

J. GREENE

DODWELL, P. C. *Visual Pattern Recognition.* Holt, Rhinehart & Wilson. 1970. Pp. 276. $9·00.

This is an unusual and, in many ways, a highly personal book. The author has concentrated on those areas of research on pattern perception and recognition with which he has had some direct contact, and, as one might expect, the direct contact has always culminated in one of his theoretical models. Professor Dodwell has now produced models for the coding of contours, for perceptual learning and development, for binocular coding, for perceptual adaptation and for attentive processes in discrimination learning. These models, bolstered by two interesting chapters on neuro-physiological data and on computer simulation, form the central core of the book.

Perhaps the most remarkable feature of this approach is what it leaves out. It is certainly a singular omission to write in this area and not to bother about, for example, the serial/parallel argument, although I myself found the omission a welcome relief. The result, however, is that Dodwell's book has to be judged almost entirely on the basis of the strengths and the weaknesses of the models which he is proposing. These are always interesting, but they certainly can be criticized.

It seems to me that these models have three weaknesses. The first, and most serious, is that none of the models produces clear predictions, which satisfactorily distinguish it from alternative hypotheses. The author himself consistently acknowledges this difficulty, but apparently can think of no way of overcoming it. He suggests for example that his model of attentive processes in discrimination learning represents a step in the right direction: but how can one know that the direction is right without the chance of meaningful evidence?

The second weakness is a common one in theoretical approaches to pattern recognition which are primarily based on discrimination learning experiments. The author consistently fudges the distinction between the perception and the recognition of patterns. When rats

and children do not learn a discrimination between oblique lines, Dodwell argues that they do not perceive the difference between these lines: but discrimination tasks involve memory from trial to trial as well as perceptual differentiation of the stimuli. It is now very clear that young children, at least, can distinguish simultaneously presented obliques, although they have difficulty in storing distinctive information about them, and, therefore, in recognizing them. This means that quite different codes are needed for simultaneous discriminations, on the one hand, and for successive discriminations of the kind that are necessary in discrimination tasks, on the other. None of Dodwell's proposed models caters for such a distinction.

The third point concerns Dodwell's treatment of perceptual development and perceptual learning. Although he quite correctly criticizes other models for avoiding the subject of perceptual change, his approach to this problem turns out to be rather an empty one. He simply adds to his contour coding model a memory which is capable of associating different perceptual events which normally tend to follow one another. It is never clear what exactly this model is set up to explain. At one point Dodwell seems to argue that it accounts for the development of perceptual constancy, but, if we are to believe Bower, the kind of constancy Dodwell is talking about is innate and does not develop. At any rate, it is not at all clear how Dodwell's model of perceptual learning can account even for the learning of simple patterns. It certainly does not explain how it is that children eventually learn to distinguish mirror image figures which they could not distinguish when they were younger.

Despite these difficulties the book is well worth reading. The author has grappled with some very difficult problems for a long time, and his book gives an honest and readable account of the results of his efforts. It is clearly written, and well diagrammed. It should be taken seriously.

P. E. BRYANT

GLADWIN, THOMAS. *East is a Big Bird.* Oxford: University Press. 1970. Pp. 241. £4·75.

Not until quite recently has any serious research been carried out into Polynesian and Micronesian methods of navigation, for all that its achievements are notable by any standards. From the island of Puluwat in the Carolines, inter-island voyages involving stretches of up to 130 nautical miles out of sight of land and across the boundary of the north equatorial and equatorial counter currents are undertaken, not only with regularity but upon the smallest pretext—to buy cigarettes for example. The navigators on these remarkable journeys, often trained in the art since childhood, use neither chart not compass, nor any other navigational artifact. Scholars, for the most part ill-equipped for investigating this kind of thing, have (with notable exceptions such as the late Sir Arthur Grimble) tended to rely on esoteric explanations derived from legend; within the last few years, however, Dr David Lewis working under the auspices of the Australian National University has broken new ground by conducting a series of controlled experiments with native navigators. Last year with the Carolinian navigator Hipour he sailed some 1150 miles from the Carolines to the Marianas, using neither charts nor instruments.

Steering in this indigenous system is for the most part by guiding stars which rise or set on an island's bearing, and a form of siderial compass is derived from the rising or setting points of thirty-two of the brightest or most easily recognized stars. The sun and such phenomena as swell direction are used as reference points during the day, and position is almost entirely determined by a remarkable system of dead reckoning. From his experience Lewis estimates that a course accuracy of $\pm 5°$ is generally attained, which is remarkably good. The skills inherent in this kind of performance are not easily acquired and the Pacific navigator is subjected to a long and rigorous training. In the Carolines, for example, the star courses governing journeys between some sixty islands stretching over perhaps two thousand miles must be committed to memory. The vast amount of information that the skilled navigator must acquire is not written down but is passed from one generation to the next by word of mouth.

The author, a visiting professor of anthropology at the University of Hawaii, underwent an apprenticeship in navigation under Puluwatan master navigators and was instructed both ashore and by demonstration afloat out of sight of land. Navigationally his book is most distinguished. It gives a clear, and well illustrated description of the Carolinian system of navigation, and the training of navigators, the role of navigation in the community, and of the construction and handling of the beautiful "flying proas" in which these voyages are made. It was, incidentally, Dr Gladwin who introduced David Lewis to Hipour.

Dr Gladwin's declared purpose in studying Carolinian navigation and the training of navigators in Puluwat, where the prestige accorded to the office is unparalleled, was to shed light on the thinking processes of the American poor who tend to become drop-outs from the educational system; his research was accordingly supported by that far seeing body, the National Institute of Mental Health. He draws an analogy between the methods used by Carolinian navigators and those employed by uneducated taxi drivers to find their way around cities like New York. He argues that both tasks require considerable skill and that they are both carried out by committing to memory in advance a large number of contingency plans so that no creative effort is needed to take decisions during the course of a journey. It is also claimed that both tasks require the ability to acquire abstractions and to respond to faint sensory cues such as the direction of a particular kind of wave picked out from other types of wave travelling in a multitude of different directions. The analogy between Carolinian navigators and New York taxi drivers is perhaps a little far fetched, particularly, as the errors made by the latter are often much in excess of the 5° that the Puluwat navigator can afford. Nevertheless, it is an interesting thought that there may be cultures that do not encourage creative thinking of the sort the psychologist calls problem solving but organize their knowledge systems in such a way as to obviate the necessity for creative thinking while at the same time allowing them to carry out highly complex tasks such as navigation. It remains to be proved that the reason why the American poor do badly at school and in conventional intelligence tests is because they come from a culture whose knowledge systems are so organized that they are not called upon to think creatively.

Although it is doubtful whether the book throws much light on the thinking processes of poor Americans, the information on the navigation system used on Puluwat is fascinating in its own right and is conveyed with delightful clarity, and the idea that a suitably organized memory may be a good substitute for creativity is an interesting one.

M. W. Richey
N. S. Sutherland

Gulick, W. L. *Hearing: Physiology and Phychophysics*. New York: Oxford University Press. 1971. Pp. viii+257. £4·50.

Plomp, R. and Smoorenburg, G. F.(Eds.). *Frequency Analysis and Periodicity Detection in Hearing*. Leiden: A. W. Sijthoff. 1970. Pp. xvi+482.

Hearing is an introductory but not quite elementary book. It renders intelligible the mechanics of the cochlea, the electrophysiology of the Organ of Corti and auditory nerve, and the anatomy of the auditory CNS. It provides some introduction to such essential psychoacoustic concepts as place and periodicity analyses of frequency, critical bands, and localization cues; in this it is preferable to the single "hearing" chapters in many introductory books. But from the experimental psychologists' point of view it has many inadequacies: the account of complex auditory phenomena is sketchy, there are relatively few references from the 1960's and the chief growth areas of psychoacoustics are ignored. While regard for the classics is welcomed, such disregard for their modification and extension can only result from an ideological commitment either to the past or to physiology to the exclusion of psychology.

Although in marked contrast to Gulick's book in several respects, *Frequency Analysis and Periodicity Detection* does not solve the problem of presenting modern psychoacoustics to a wider audience either. An intelligently planned and edited symposium, it prefaces five main sections with review papers which will be of use to non-specialists, although they

assume some familiarity with acoustical terminology since the symposium was quite specialized. The experimental papers are of a very high standard, as is the subsequent discussion of each. Experimental issues such as jitter detection, after-effects, central masking, diplacusis, and the duration-dependence of frequency-sensitivity are thrashed out by the respective experts. Fourcin's crucial experiments on pitch generation by interaural delays of white noise are now conveniently available. Plomp's lucid exposition of experimental definitions of timbre can be taken along with the official review papers which include Schouten and Ritsma on periodicity pitch, Zwicker on frequency analysis and masking, as well as physiological summaries of mechanical, single fibre, and central functions in audition. This book marks the solid reinstatement of the periodicity principle ("volley theory" of 20 years ago) as an important aspect of the hearing process and depicts progress towards descriptions that integrate the contribution of place and periodicity mechanisms to auditory sensation. No sensory psychologist should be without it.

M. P. HAGGARD

JAFFE, JOSEPH AND FELDSTEIN, STANLEY. *Rhythms of Dialogue.* New York: Academic Press. 1970. Pp. 156. £3·95

This monograph by two psychiatrists of multidisciplinary orientation will be of some relevance to behavioural, social and communication scientists, particularly those interested in apparatus and equipment; in studying and understanding conversation; or in mathematical modelling.

The equipment consists of a completely automated system of interaction chronography which can be applied to two-person verbal interactions occurring at close range in a face-to-face situation where the only equipment apparent to the speakers consists of two ordinary microphones. The authors describe sequences of sound and silence in terms of only five categories—vocalization (by identified speaker); pause (by speaker); switching pause (silence preceding switch of speaker); speaker switch; simultaneous speech (from both speakers)—they analyse what is described as "conversational rhythm" by inputting live or tape-recorded material to an on-line digital computer by means of a two channel speech detector and analogue to digital converter. The special feature of the converter is a network which electronically cancels the "spill" of each speaker's voice in to the other's microphone.

Approximately 600 conversations were analysed. The person studying conversation will encounter several interesting, if piecemeal, findings. The distributions of the classification categories we found, in general, to be exponential. The time patterns of an individual's conversational style remained markedly stable during the course of a conversation and consistent from one conversation to another with the same partner. There was a tendency for partners in a conversation to match each other's silences, but this interactional effect did not occur with speech categories. Insofar as sex differences appear to have occurred, they did so not as clear main effects but as parts of difficult to interpret statistical interactions. The total absence of visual-gestural information did not markedly affect vocalization but did reduce the length of pauses and of simultaneous speech.

These and other findings, though very limited can, be related to, or contrasted with, other evidence regarding conversation. Since these findings emerged relatively early in the research programme, one might have hoped that subsequent empirical work would have been carried out to extend them or that the authors would have attempted to construct a "theory" to explain them to and generate more research. Instead, they opted to construct mathematical models, four in all, to describe their findings. The models are all Markov chain models, the simplest being a four-state "single-source" (i.e. monologue) model and the most complex, and most satisfactory, being a six-state model which assumes two separate speakers. The authors rightly make the point that these models potentially do more than redescribe the previously presented data. They could be used to predict new sequences of behaviour, to reinterpret research performed by others and to generate new research to examine the implicit psychological assumptions of the models. The authors, however, do none of these, and one is left with the feeling that a mathematical model

of very little still amounts to very little, and four such models amount to not much more.

From the point of view of someone who feels that the single most important gap in the study of language is the investigation of conversation, the payoff from this monograph is disappointing. If the book is intended as a progress report then one can still look forward to the equipment and the models eventually producing evidence and insight. If the book is intended as a final report on a research programme, then it seems as if technology has scored another victory over understanding.

C. FRASER

Publications Received

BROADBENT, D. E. *Decision and Stress.* London: Academic Press. 1971. Pp. 536. £7·25.

COOMBS, C. H., DAWES, R. M. AND TVERSKY, A. *Mathematical Psychology, an Elementary Introduction.* Prentice-Hall. 1970. Pp. 419. £5·50.

COOPER, C. L. AND MANGHAM, I. L. (Eds). *T-groups, a Survey of Research.* John Wiley. 1971. Pp. 283. £4·50.

DAVIS, P. R. *Performance under Sub-optimal Conditions.* London: Taylor & Francis 1971. Pp. 97. £2·25.

FELDMAN, M. P. *Psychology in the Industrial Development.* London: Butterworth. 1971. Pp. 128. £2·50.

JONES, C. AND SABATER PI, J. *Bibliotheca Primatologica, No.* 13. Basel: S. Karger Ag. 1971. Pp. 96. £2·65.

KAPLAN, M. L., COLARELLI, N. J., GROSS, R. B., LEVENTHAL, D. AND SIEGEL, S. M. *The Structural Approach in Psychological Testing.* New York: Pergamon Press. 1971. Pp. 206. £3·50.

LINDGREN, H. C. AND BYRNE, D. *Psychology, an Introduction to a Behavioural Science.* New York: John Wiley. 1971. Pp. 759. £5·50.

PLOMP, R. AND SMOORENBURG, G. F. (Eds.). *Frequency Analysis and Periodicity Detection in Hearing.* Leiden: Sijthoff. 1970. Pp. 482. Fl.60.

RADNER, M. AND WINOKUR, S. (Eds.). *Analyses of Theories and Methods of Physics and Psychology.* London: Oxford University Press. 1971. Pp. 441. £6·00.

RITCHIE RUSSELL, W. *The Traumatic Amnesias.* Oxford University Press. 1971. Pp 84. £2·00.

SCOTT MOSS, C. *Dreams, Images and Fantasy: a Semantic Casebook.* Chicago: University of Illinois Press. 1971. Pp. 302. £4·10.

SHAW, M. E. *Group Dynamics, the Psychology of Small Group Behavior.* New York: McGraw-Hill. 1971. Pp. 414. £5·25.

SINGLETON, W. T., FOX, J. G. AND WHITFIELD, D. (Eds.). *Measurement of Man at Work: an Appraisal of Physiological and Psychological Criteria in Man–Machine Systems.* London: Taylor & Francis. 1971. Pp. 267. £6·00.

SOLING, H-D. AND WILLMS, B. (Eds.). *Regulation of Gluconeogenesis.* Stuttgart: Georg Thieme Verlag. 1971. Pp. 357. DM.46.

VERNON, P. E. *The Structure of Human Abilities.* London: Methuen. 1971. Pp. 208. £0·90.

ZUSNE, L. *Visual Perception of Form.* New York: Academic Press. 1970. Pp. 547. £9·10.

CONTENTS

Printed by W Heffer & Sons Ltd Cambridge England

This issue completes Volume 23

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Volume 23 Part 4 November 1971

PUBLISHED FOR THE EXPERIMENTAL PSYCHOLOGY SOCI
BY ACADEMIC PRESS LONDON AND NEW YORK

NOTES FOR AUTHORS

Only original papers are accepted for publication. All manuscripts should be submitted in *duplicate* to *Professor N. S. Sutherland, Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex, BN1 9QY, England.* One copy must be a *typewritten original* with double spacing and wide margins. Figures should be drawn to professional standards in Indian ink on tracing linen or Bristol Board; the original drawing (plus one copy) should be submitted. Allowance must be made for a reduction in size when the drawing is reproduced (the width of lines as reproduced should be between 0·25 and 0·33 mm). Explanations must be placed in the figure legend; the drawing should contain as little lettering as possible and lettering should be inserted in *blue* pencil or on a tracing paper to fit over the drawing. Tables and figure legends should be typed on separate sheets of paper. Where possible related diagrams should be grouped together to form a single figure.

Intending contributors should pay attention to the *Journal*'s conventions, and in particular to the form of headings, and sub-headings used, and references (in the Reference list titles of journals will in future be printed in full without abbreviation). The *Journal* will use the *Système Internationale* for all measurements (if terms like inches are used, metric equivalents must be quoted). The use of standard abbreviations in the text is permissible (if in doubt, write term in full the first time it is used and give abbreviation in brackets); please note that the *Journal* does not use S for subject.

An abstract must be included. Although abstracts should be as tersely written as possible, they should summarize all the main experimental findings. The length of abstracts will vary with the length and complexity of the experiments reported but should usually be between 50 and 400 words.

An initial decision on a paper will normally be reached within 2 months of receipt. Where only minor revisions are requested, date of publication will be determined by date of receipt of the paper provided a satisfactory revision is received within 1 month of the request for a revision being made. The delay between final acceptance of a paper and its publication will normally be between 3 and 6 months. The date on which the manuscript is received in its final version (apart from minor alterations—see above) is printed.

Contributors receive 25 copies of their paper free. If they require further copies they should apply to the Publishers at the time of returning proofs.

The Quarterly Journal of Experimental Psychology is published for The Experimental Psychology Society by Academic Press, London and New York. Volume 23, 1971 (4 issues): subscription rate £6.50 plus £0.40 postage inland; $16.00 plus $1.00 postage abroad. Subscription orders should be addressed to Academic Press Inc. (London) Limited, 24–28 Oval Road, London NW1 7DX with the exception of those originating in the U.S.A., Canada, Central America and South America; these should be sent to Academic Press Inc., 111, Fifth Avenue, New York, New York 10003. Application to mail at second-class postage rates at Jamaica, N.Y. 11431.

Undergraduates and graduates of less than three years' standing may obtain the journal at the reduced price of £3.60. They should apply initially to the Hon. Secretary, Dr David Legge, Department of Psychology, City of London Polytechnic, Central House, Whitechapel High Street, London, E1 7PF. They should not send any remittance, but enclose suitable evidence of their student status or date of graduation, such as a letter from a university teacher or copy of a graduation list.

Quarterly Journal of Experimental Psychology (1971) **23**, 359–366

BLOCKING AS A FUNCTION OF NOVELTY OF CS AND PREDICTABILITY OF UCS

N. J. MACKINTOSH
Dalhousie University
AND C. TURNER
University of Edinburgh

Kamin (1969) reported that after rats had been trained with one CS predicting shock, they learned virtually nothing about a second CS added to the first to form a compound—provided that the shock remained unchanged. If, however, the shock was either increased or omitted at the same time that the new CS was added, it acquired significant excitatory or inhibitory strength. Both these results were confirmed in the present experiment. In addition, however, it was found that this excitatory or inhibitory conditioning to the added element could itself be blocked if the shock remained unchanged for the first four trials after the addition of the new element. This finding implies that blocking is not due to any limitation on attentional or associative capacity, but rather occurs when subjects learn that a particular stimulus is redundant, i.e. that it predicts no change in reinforcement.

Introduction

Kamin (1968, 1969), studying the suppression of appetitively reinforced lever pressing in rats by the presentation of a CS signalling shock, found that prior training on one element of a compound CS blocked conditioning to the other element. After 16 conditioning trials to white noise, eight trials with a compound noise and light CS resulted in no more suppression to the light than was shown by animals for whom the light had never been paired with shock.

One interpretation of this result is in terms of a theory of selective attention (e.g. Sutherland and Mackintosh, 1971). Prior training with the noise would be said to ensure a high level of attention to this stimulus, with a consequent reduction in the probability of attending to any other stimulus presented at the same time as the noise. Kamin himself suggested a different explanation, arguing that no learning about the added element occurred on compound trials, because the occurrence of shock was already fully predictable on the basis of the previously trained element (Kamin, 1969). The assumption is that only unexpected reinforcers are reinforcing. This idea was formalized by Wagner (1969) and elaborated by Rescorla and Wagner (1971) into a systematic account of conditioning. Increments in excitatory strength to a given stimulus consequent upon its association with a reinforcing event are assumed to be a decreasing function of the current level of excitation accruing to the entire compound of which the stimulus forms a part. When the excitatory strength of the compound has reached the asymptote

of conditioning supported by the particular reinforcer used, then no further increments in excitatory strength can occur to any newly added element.

This latter account has one advantage over an attentional analysis, for it provides a simple and straightforward explanation of two cases reported by Kamin (1969) in which prior training with a noise CS did not block conditioning to the light on compound trials. In the first case, shock was omitted on compound trials, and there was evidence that the light acquired significant inhibitory strength. In the second case, the intensity of the shock was increased on compound trials, and the light acquired significant excitatory strength. Kamin (1969) argued that both the omission of shock and the increase in shock intensity were unexpected, and therefore effective reinforcing events. Rescorla and Wagner's (1971) theory accounts for the inhibitory learning because increments in inhibition to an element associated with non-reinforcement are an increasing (instead of decreasing) function of the excitatory strength of the compound of which it forms a part; and they account for the excitatory learning because the asymptote of excitatory strength is an increasing function of the intensity of the UCS used, and the introduction of a stronger shock means that there is room for further excitatory conditioning to occur.

It might be possible to explain these two instances of "unblocking" in terms of the attentional analysis outlined above, by suggesting that the change in reinforcement disrupts attention to the noise and hence permits learning about the light. This somewhat *ad hoc* account, however, is probably not worth pursuing, since the results to be reported below suggest a quite different approach. The present study was prompted by the observation that in the two cases where conditioning occurs to the light, its introduction into the experimental situation is exactly correlated with a change in reinforcement. Perhaps a stimulus only acquires signalling properties (either excitatory or inhibitory) when its introduction signals such a change. In the standard blocking experiment, of course, the introduction of the added element signals no change in reinforcement. The implication is that blocking is a consequence of subjects having learned that the added element predicts no change, i.e. that it is redundant. From this it follows that one might block learning in Kamin's two unblocking experiments by teaching subjects that the light, on its first introduction, is redundant. This could be achieved by scheduling a small number of trials with the noise-light compound paired with the original UCS before the compound is paired with the changed UCS.

Method

Subjects and apparatus

The subjects were 48 experimentally naive, male hooded rats, obtained from Les Elevages Medico, Montreal, and weighing 250–300 g on their arrival in the laboratory. They were maintained at 85% of their *ad lib* weights. The apparatus consisted of two Grason-Stadler one-lever rat chambers modified by the addition of an overhead 40 W light which was diffused through a white Plexiglass ceiling. The boxes were enclosed in sound-attenuating shells and automatic programming equipment was located in an adjoining room. The stimuli used in the experiment were as follows: N = a 70 dB white noise; L = the turning on of the overhead light; T = an intermittent (0·5 sec on, 0·5 sec off) 70 dB, 1800 Hz tone.

Procedure

Subjects were placed in the apparatus with 45 mg Noyes pellets in the food tray, and left until they had pressed the lever 50 times for 50 more pellets. On the next 2 days they were placed on a variable interval schedule of food reinforcement which was gradually lengthened to VI 90 sec. For the next 6 days, they received daily one hour sessions of VI 90 sec training.

After this preliminary training, subjects received four pre-test trials to each of the stimuli used in subsequent training (all subjects received *N* and *L* trials, and some received *T* trials as well). Pre-test sessions contained four stimulus presentations at 15-min intervals, with each stimulus presented for 1 min. Experimental training began on the day following the final pre-test session. In each 1 hr session, subjects received four conditioning trials on which a 1 min CS was followed by a 1 mA 0·5 sec scrambled grid shock. Trials occurred at approximately 15-min intervals starting with the 13th min of each session. Occasional subjects who showed an overall decline in lever pressing were left in the apparatus at the end of their one hr session until responding recovered.

Experimental design

The design of the entire experiment is shown in Table I. There were two main pairs of experimental groups, and a control condition. The first pair of experimental groups comprised the inhibitory condition. Group Inh-C (Inhibitory Control) was designed to replicate Kamin's unblocking result. After initial training in Stage 1 to N and T (three N trials and one T trial per day), they continued in Stage 2 to receive reinforced T trials alternated with non-reinforced presentations of the NL compound. The test of L as a conditioned inhibitor consisted in comparing (over a series of non-reinforced test trials) suppression to the *TL* compound with suppression to *T* alone. Group Inh-E (Inhibitory Experimental) was treated in exactly the same way, except that they received four *NL* trials with the original 1 mA shock at the end of Stage 1, i.e. on the day before inhibitory training to *NL*.

TABLE I

Experimental design

Groups	Stage 1		Stage 2	Test
Ihn-C	24*N*, 8*T*→Shock	———	10*NL*→No Shock, 10*T*→Shock	4*TL*, 4*T*
Inh-E	24*N*, 8*T*→Shock	4*NL*→Shock		
Exc-C	24*N*→Shock	———	4*NL*→SHOCK	4*L*
Exc-E	24*N*→Shock	4*NL*→Shock		
Blocking Controls	24*N*, 8*T*→Shock, *or* 24*N*→Shock		——— 4*NL*→Shock	4*L*

N = Noise; *T* = Tone; *L* = Light; Shock = 1 mA shock; SHOCK = 4 mA shock.

The second pair of experimental groups comprised the excitatory condition. Group Exc-C was designed to replicate Kamin's results. After 24 trials with *N* paired with a 1 mA shock in Stage 1, they received four trials on which the *NL* compound was paired with a 4 mA shock in Stage 2, and were finally tested (in extinction) for suppression to *L*. Group Exc-E received identical treatment, with the addition, at the end of Stage 1, of four trials on which *NL* signalled the original 1 mA shock.

A final pair of control groups completed the experimental design. These groups were used to ensure that the stimuli and general experimental procedures we employed were

such as to reproduce Kamin's basic blocking result. The question was whether animals receiving 24 N trials followed by 4 NL trials all with the same 1 mA shock would show any suppression to L: their performance on test trials to L was compared with that of subjects receiving 24 N trials alone. Two control conditions were run: for one, Stage 1 training was the same as for the inhibitory condition (8 T trials interspersed with 24 N trials), for the second, Stage 1 duplicated the excitatory condition (24 N trials only).

Results

In order to allow for individual differences in rate of responding, the effects of CS presentations on responding were expressed as suppression ratios of the form $a/(a+b)$ where $a =$ the number of responses during the CS, and $b =$ the number of responses in the 1 min interval preceding the CS. A ratio of 0·00, therefore, indicates no responding during the CS, i.e. complete suppression, while a ratio of 0·50 indicates an equal rate of responding before and during the CS, i.e. no suppression. Except where noted below, responses during pre-CS periods and during CSs were cumulated for all trials with that CS in a given session, and a single suppression ratio was calculated for each subject for each session.

During Stage 1, in all conditions of the experiment, suppression was essentially complete to both N and (where applicable) T after 3 days of training. On the last day of Stage 1, group mean suppression ratios ranged from 0·00 to 0·03. In Stage 2, in the inhibitory condition, Group Inh-C showed a rapid loss of suppression on non-reinforced NL trials, whereas Group Inh-E maintained a greater degree of suppression. The mean suppression to NL over the entire five days of Stage 2 was 0·29 in Group Inh-C and 0·11 in Group Inh-E. This difference was significant when evaluated by a Mann-Whitney U-test ($U = 13$, $P = 0{\cdot}025$). By the final day of Stage 2, however, the difference (0·35 for Group Inh-C, and 0·24 for Group Inh-E) had fallen short of significance ($U = 21$, $P = 0{\cdot}139$). The test results are shown in Figure 1. As can be seen, Group Inh-C showed much less suppression to TL than to T, while Group Inh-E showed equally strong suppression to both stimuli. Only in Group Inh-C, therefore, had non-reinforced NL trials established L as a strong conditioned inhibitor. Wilcoxon

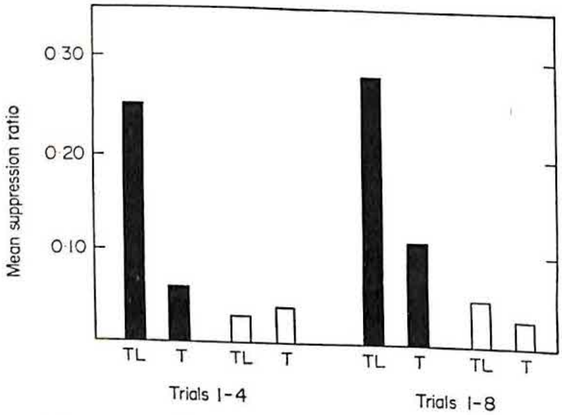

FIGURE 1. Test results for inhibitory groups.
■, Group Inh-C; □, Group Inh-E.

matched-pairs tests showed that both on the first test day and on the two test days combined, Group Inh-C was significantly less suppressed to *TL* than to *T* (in each case, $T = 1$, $P < 0{\cdot}02$), while neither difference approached significance in Group Inh-E ($T = 16$ and 9; $P > 0{\cdot}10$). Finally, each subject was assigned a difference score between suppression ratio to *T* and suppression ratio to *TL*. Mann-Whitney *U*-tests on these scores revealed a significant difference between groups both on Day 1 ($U = 9$, $P = 0{\cdot}007$) and on Days 1 and 2 combined ($U = 10$, $P = 0{\cdot}010$).

After prior conditioning to *N*, therefore, the presentation of an *NL* compound which from the outset signalled non-reinforcement resulted in significant inhibitory conditioning to *L*. But four reinforced *NL* trials prior to inhibitory training blocked this inhibitory conditioning.

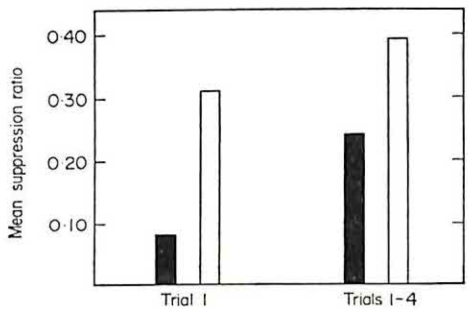

FIGURE 2. Test results for excitatory groups.
■, Group Exc-C; □, Group Exc-E.

In the excitatory condition, learning about *L* during *NL* compound trials should have been evidenced by significant suppression to *L* during test trials. Figure 2 shows these results: Group Exc-C showed substantial suppression to *L* (thus confirming Kamin's results); Group Exc-E, however, showed considerably less suppression. Both on Trial 1 and on Trials 1–4 the difference between groups was highly significant (Trial 1: $U = 6$, $P = 0{\cdot}002$; Trials 1–4: $U = 11$, $P = 0{\cdot}014$). Once again, four trials of exposure to the *NL* compound with unchanged reinforcement significantly impaired conditioning to *L* when reinforcement did change.

The results of the two main control groups were similar, and were pooled for statistical analysis. The eight subjects receiving four *NL* trials following 24 *N* trials (with or without interspersed *T* trials) showed no greater suppression to *L* than did the eight subjects given no *NL* trials. On Trial 1, the suppression ratios were 0·39 for the group exposed to *NL*, and 0·40 for the group receiving *N* trials only; over Trials 1–4, the suppression ratios were 0·45 and 0·44 respectively. Neither of these differences approached significance (in both cases, $P > 0{\cdot}40$). On this measure, therefore, the administration of four reinforced *NL* trials produced no conditioning to *L* at all—although, of course, these trials had substantial effects on *subsequent* conditioning to *L* in the two main experimental conditions.

In fact, Kamin has presented some evidence to suggest that the use of a more sensitive relearning measure may reveal a very small amount of learning about *L* during *NL* compound trials. It is possible, therefore, that such marginal learning

occurred in the present experiment. There are several reasons, however, why this possibility cannot seriously affect the conclusions that may be drawn from the present results. First, the test procedures used to reveal the effects of four reinforced *NL* trials on ***subsequent*** conditioning to *L* in the main experimental conditions were exactly the same as those which revealed no conditioning to *L* *during* those four trials: in both cases non-reinforced test trials rather than the more sensitive relearning measures were used. Second, any conditioning that may have occurred to *L* during these four *NL* trials cannot easily explain the results of both the Inhibitory and Excitatory conditions. If such conditioning occurred, then *L* would have had some excitatory strength at the outset of Stage 2. This might have retarded the conditioning of inhibition to *L* in the Inhibitory condition, and thus have been partially responsible for the outcome of this portion of the experiment. But exactly the same argument implies that, in the Excitatory condition, subjects receiving the four reinforced *NL* trials in Stage 1 should have shown *more* suppression to *L* than those that received *NL* training only in Stage 2. Finally, Kamin (1969) has reported that any learning that occurs about the added element is confined to the first compound trial, and that on this trial, suppression is significantly disrupted: the implication is that suppression is momentarily below asymptote, and therefore that there is room for further conditioning to occur. Kamin has presented extensive data on this transient disruption of suppression: in his equivalent of our Group Inh-C, for example, the median suppression ratio on the first compound trial (before the first non-reinforcement could have taken effect) was between 0·15 and 0·20. Whatever the cause of this disruption in Kamin's procedure, it did not occur in the present experiment. As noted above, our normal procedure was to calculate a single suppression ratio for each subject for all trials with a given CS in each session. We did, however, collect data separately from the first *NL* trial for all subjects in Group Inh-E, since it was in this group that excitatory conditioning to *L*, if it had occurred, would have counteracted the development of inhibition. Seven of the eight subjects in this group had suppression ratios of 0·00 on this first compound trial; the eighth had a suppression ratio of 0·11. Of ten further subjects for whom data were collected for the first compound trial, eight had suppression ratios of 0·00. If the small amount of conditioning to the added element observed by Kamin, is, as he suggests, due to the disruption of suppression on this trial, the only reasonable conclusion is that with the procedures used here very much less conditioning can have occurred. If this is so, then it becomes even less plausible to suggest that such marginal amounts of conditioning to *L* during reinforced *NL* trials could have had the rather substantial effects on Stage 2 conditioning that were observed.

Discussion

The results of the present experiment seem unequivocal. After suppression has been completely conditioned to *N*, four trials of exposure to an *NL* compound and unchanged reinforcement markedly attenuate subsequent conditioning to *L* when the *NL* compound signals a change in reinforcement. As it happens, in the inhibitory condition there was no evidence of any inhibitory conditioning to *L*

in Group Inh-E: with this measure, therefore, blocking was complete. A similar conclusion is not possible in the excitatory condition since no control group was run to measure the amount of unconditioned suppression to L. It is not, however, a matter of great moment whether blocking was or was not complete, since the magnitude of any blocking effect will presumably be a function of a variety of parameters (including the number of trials in Stages 1 and 2, the number of reinforced NL trials for the two experimental groups, the relative intensities of N and L, and the magnitude of the change in the UCS). The important point is that a small number of NL trials with unchanged reinforcement significantly impairs later learning about L. The obvious implication is that this subsequent conditioning is blocked because subjects have learned that L is redundant, that is to say, have learned that its initial presentation signalled no change in reinforcing events. If this is so, then it suggests a somewhat different interpretation of Kamin's basic blocking effect: the reason why no conditioning occurs to L when NL signals unchanged reinforcement is that subjects learn that L is redundant and therefore "ignore" it.

These conclusions are at odds with both major interpretations of blocking discussed in the introduction. Subjects do not fail to learn about the light because they are fully occupied attending to the noise and cannot therefore attend to the light. On the contrary, they *do* attend to the light and thereby learn that it is redundant. Nor does Rescorla and Wagner's (1971) analysis fare any better. The absence of blocking when reinforcement is either increased or decreased cannot be due to the fact that these changes establish new asymptotes, and hence permit further (excitatory or inhibitory) conditioning, for, as we have argued, the four redundant NL trials cannot have had the required effects on associative strength to block such conditioning. But if the absence of blocking is not due to changes in asymptote, then perhaps the occurrence of blocking is not due to an unchanging asymptote. Certainly, if we accept that the blocking of subsequent conditioning to the light observed in Groups Inh-E and Exc-E should be assimilated to the standard blocking effect, then an account that talks of learned redundancy seems much more appropriate. This notion cannot be represented in Rescorla and Wagner's model by any parameters relating to asymptotes of conditioning. The only relevant parameter in their model is α—a learning rate parameter specific to each stimulus: if redundant training retards subsequent conditioning to L, then this could be represented as a decrease in α_L.

It is tempting to say that this parameter—if permitted to change in this manner—represents something closely related to the concept of attention. But this cannot herald a return to the simple attentional analysis already discussed. If we are to talk about attention at all here, the notion needs considerable elaboration. Instead of saying that subjects fail to attend to the light on compound trials because they are attending to the noise, we seem required to say that they fail to attend to the light on later compound trials because on earlier trials they have learned that it is redundant. It is not so much a failure of attention as an active suppression of attention to uninformative stimuli.

There is a parallel with another set of results. Repeated, non-reinforced presentation of a stimulus retards subsequent conditioning to that stimulus: this is

the phenomenon of "latent inhibition" studied by Lubow and Moore (1959). This result is not due to the conditioning of inhibition to the stimulus during non-reinforced presentation, for the procedure retards subsequent inhibitory conditioning as well as excitatory conditioning (Rescorla, 1971). It seems reasonable to argue that non-reinforced presentation of a stimulus prior to conditioning allows the subject to learn that that stimulus signals no change in reinforcement, because it is uninformative or redundant in exactly the same way as was the light during the present blocking experiments. The implication is that whenever an environmental change occurs (a stimulus is presented) without further changes of significance to the subject, then attention to that stimulus, initially aroused perhaps by its novelty, is actively inhibited. The suggestion is that understanding of blocking in conditioned suppression experiments will be furthered by investigations of latent inhibition or habituation.

This research was supported by Grant APA 259 from the National Research Council of Canada.

References

KAMIN, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. JONES (Ed.), *Miami Symposium on the Prediction of Behavior: Aversive Stimulation*, pp. 9–33. Miami: University of Miami Press.

KAMIN, L. J. (1969). Predictability, surprise, attention and conditioning. In B. CAMPBELL and R. CHURCH (Eds.), *Punishment and Aversive Behavior*, pp. 279–96. New York: Appleton-Century-Crofts.

LUBOW, R. E. and MOORE, A. U. (1959). Latent inhibition: The effect of nonreinforced preexposure to the conditioned stimulus. *Journal of Comparative and Physiological Psychology*, **52**, 415–19.

RESCORLA, R. A. (1971). Summation and retardation tests of latent inhibition. *Journal of Comparative and Physiological Psychology*, **75**, 77–81.

RESCORLA, R. A. and WAGNER, A. R. (1971). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. BLACK and W. F. PROKASY (Eds.), *Classical Conditioning II*. New York: Appleton-Century-Crofts. In press.

SUTHERLAND, N. S. and MACKINTOSH, N. J. (1971): *Mechanisms of Animal Discrimination Learning*. New York: Academic Press.

WAGNER, A. R. (1969). Stimulus validity and stimulus selection in associative learning. In N. J. MACKINTOSH and W. K. HONIG (Eds.), *Fundamental Issues in Associative Learning*, pp. 90–122. Halifax: Dalhousie University Press.

Received 20 *August*, 1971

Quarterly Journal of Experimental Psychology (1971) 23, 367–372

ORIENTATION OF STIMULI AND BINOCULAR DISPARITY CODING

J. P. FRISBY AND BARBARA ROTH

Department of Psychology, University of Sheffield

An experiment is described which attempts to relate physiological work on disparity coding in the cat to a psychophysical situation using human subjects and Julesz stereograms composed of small line elements. It was found that depth perception occurred only if matching disparate lines in each stereogram shared a similar orientation. Depth began to deteriorate if an orientation difference exceeded 10° and it was extinguished at about 60°. The results are interpreted as supporting the hypothesis that shape-sensitive disparity detectors of the kind found in the cat exist also in man.

Introduction

Some recent neurophysiological investigations have thrown light on binocular depth processing in the cat (Barlow, Blakemore and Pettigrew, 1967; Blakemore, 1970; Pettigrew, 1965). Investigating binocularly driven striate single units, these workers have discovered a disparity coding system which extracts from each monocular field contours of like orientation before assessing their disparity. The present experiment sought to find out using psychophysical measures if a system similar to this existed in man. Subjects were shown stereograms of the kind used by Julesz (1961). Each was composed of many small black and grey line elements arranged in random orientations on a white ground, and disparate square areas were created in the usual Julesz fashion. The experiment investigated the effect of causing the lines composing the disparate square in one stereogram to have different orientations from their matching lines in the other stereogram. It was predicted that if monocular pattern analysis of the kind found in the cat mediated the disparity assessment, then when each matching pair of disparate line elements possessed an orientation difference great enough to disrupt markedly disparity cell activity (around 10°), the perception of the square in depth should deteriorate and ultimately, for extreme orientation differences, vanish.

It should be noted that this prediction goes beyond simply specifying that depth should be "affected" by the manipulation of line orientation in the two fields, a result which would also be expected from the point-for-point disparity model advanced by Julesz (1961, 1965, 1967). Julesz has interpreted many of his random-dot stereograms as showing that monocular pattern analysis of each field is not a *necessary* precursor of disparity assessment (although it may well occur in some circumstances) and that a point-for-point comparison of the two fields can alone be sufficient. This conclusion has been supported by the results of Kaufman and Pitblado (1965) on letter matrix stimuli, and of Mitchell (1969), on tachistoscopically presented diplopic single line stimuli. Now, a point-for-point model would

also predict that the manipulation of line orientation used here would affect the perception of depth because as the lines of the disparate square are twisted with respect to one another, so point-for-point similarities between the two fields are diminished. Given then that either kind of system predicts the existence of an orientation effect, if the hypothesis that cat-type disparity detectors exist in man is to be effectively confirmed by the results of the present experiment, it must be on the basis of how well its predictions about the *detailed* nature of the orientation effect are met. The same basic consideration applies to the point-for-point account also, of course, but this model has not yet been sufficiently well developed, as far as we know, to permit any detailed predictions of the required kind to be drawn.

Method

Apparatus

A 2-ft square of thick white card, strengthened with a wooden frame, was used as a base for mounting 2,500 small line elements. These elements were prepared by cutting up welded chicken wire to obtain T-pieces. Half of them were painted black and half left silver-grey. The stems of the elements were then inserted at 15 mm intervals into the card so that on one side the heads of the T-pieces formed 10×1 mm bars arranged in regular rows and columns. Black and grey elements were inserted randomly to give both random brightness and random orientation distributions. The overall appearance was of a mass of small black and silver lines lying in all directions, as macropatternless subjectively as any of Julesz's stereograms. A black and white slide of this "standard" stimulus was taken with a Polaroid camera using 146-L half-tone film. A central square of 100 elements was then shifted laterally by one column, the vacant area so created being filled randomly with the displaced elements. Two precautions were taken to ensure that line orientations were maintained at their original values following this shift: (a) peripheral elements not forming the square were glued in position from the outset, and (b) the first slide taken was projected on to the cardboard mounting so that the orientations of the moved elements could be made to replicate their partners in the standard stimulus. A second slide was then prepared of this 0° orientation difference (0° θ) condition. Then eight more stimuli were arranged and photographed, each possessing a different amount of orientation difference (see the abscissa of Fig. 1 for the θ values chosen). Orientation changes were effected by twisting the stem of each T-piece composing the disparate square, making the lines vary in orientation about their midpoints. Alignment of the lines was made reasonably accurate (*c.* $\pm 1°$) by fixing on the stem of each manipulated T-piece a small plastic block which served to keep the lines firmly in position and which made adjustments fairly easy. Although the lines as photographed varied randomly in their original orientation, the plastic blocks on the back of the cardboard mounting were all arranged in one direction, which made it relatively easy to see that they were all twisted by the same amount. A stereoscopic Polaroid projector was used to present the stimuli to subjects standing about 3·5 m from a screen. The visual angle of each element was thus about 0·2° and the size of the disparity was about 0·3°.

Subjects

Five women and three men aged 20–30 years acted as subjects. All but one were naive undergraduate students of psychology; the remaining subject was J.P.F. and B.R. did all the testing.

Design

The experimental stimuli proper were preceded by a stereogram taken from Julesz (1965) which possessed both an area which floated forward and an area which receded in depth. This was used to familiarize the subjects with the situation and to screen for "depth blindness" (Julesz, 1965); one potential subject was in fact eliminated on this basis. The experi-

mental series comprised ten conditions in all and each subject saw them in a different random order. Nine were stereo-pairs presenting the various θ conditions, each using the "standard" stimulus and one of the θ series. The tenth was a "control" condition made up of the 0° and 90° pair: its function will become clear later.

Procedure

The following instructions were first read by each subject: "When I tell you to, put on the glasses and look at the slide on the screen. When you feel ready, describe anything you see which you feel is worthy of comment. (Sometimes any effect may take a little time to build up.) When you have looked at a few of the slides, you may find that you begin to expect certain effects; please try not to be influenced by any development of 'set'." The room was then darkened and the stimuli were presented. Subjects were encouraged to rest at any point if they wanted to since the task appeared to be a taxing one and required concentration. Subjects who had spectacles wore the experimental polaroid glasses in front of them. Attempts were made to allow sufficient time for the build-up of any effect: it is well-known that Julesz stereograms are often not perceived in depth immediately by naive viewers (Julesz, 1961). The longest viewing time allowed was *c.* 2 min, but the usual time was much shorter and only *c.* 10 sec in some conditions. No data were, however, systematically recorded on this. Although prompting was avoided, some amplification of responses was obtained by asking a subject to compare the elements in the central square, if such an area had been reported, with those in the surrounding border in terms of size and brightness. Subjects were also asked to report on the effects, if any, of fixating a point a few units to one side of the square, and to say whether or not the central square moved as a unit when they moved their heads from side to side. It was at all times stressed that there were no "correct" responses and that people varied greatly in how they perceived these stimuli. Sometimes some stimuli were presented for a second time if the subject's responses seemed to require clarification. Each stimulus pair was viewed through spectacles "upside-down" as well as the "right way up", in a random order. This had the effect of reversing the disparity and thus the nature of the depth impression, if any. The subjects' comments were written down by the experimenter as far as possible verbatim.

Results

The verbatim responses were rated by B.R. and by two independent non-psychologists working from unlabelled copies of the original responses. A 4-point scale (based upon a preliminary inspection of the data) was used as follows: (1) a clear impression of depth; (2) an impression of depth but not of maximum clarity; (3) an ambiguous impression: either oscillating depth/no depth, or else great difficulty in saying either way; (4) no impression of depth: sometimes a vague square seen but not in depth.

This retrospective assessment was preferred to the more common self-rating technique using a prearranged scale because it seemed less likely to bias the subjects or to channel their responses into preconceived but inappropriate categories.

Ratings given to forward and backward depth responses were pooled because subjects never saw depth in the inappropriate direction and because the ratings given to each direction were so rarely different. Figure 1 shows the group mean ratings for the various θ conditions. The judgements of the three raters about each subject's responses were combined in forming this curve, a justifiable procedure in view of the high inter-rater agreement (r values between their group means were 0·927, 0·932 and 0·951). The data from J.P.F. were included as inspection indicated that they were little different from those of the other, naive subjects.

The deterioration in depth as θ increases is highly significant (Friedman's $\chi^2 = 51{\cdot}6$, $df = 8$, $P < 0{\cdot}001$).

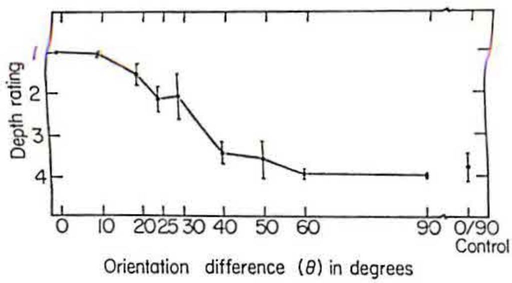

FIGURE 1. The effect of orientation difference on perceived depth. The vertical lines represent 95% confidence intervals ($N = 8$).

Discussion

The data clearly support the hypothesis that disparity detectors of the kind found in the cat exist also in man. Figure 1 suggests that when θ exceeds c. 10° the perception of depth begins to deteriorate, becoming finally extinguished when $\theta \approx 60°$. This finding of orientational selectivity in the human visual system is in keeping with much recent work which, like the present experiment, has been motivated by a desire to relate neurophysiological advances to psychophysical studies of perception (e.g. Blakemore and Campbell, 1969; Campbell and Kulikowski, 1966; Campbell and Maffei, 1970; Frisby, 1971). Indeed, the S-shaped θ effect depicted in Figure 1 is very similar to that found by Frisby for the effect of θ on optimal apparent movement. He suggested that "complex cells" coded the "moving" line stimulus and that later stages of processing examined the output of these units in such a way that if a certain "criterion rate of firing" was absent, then the chances of perceiving optimal movement were much reduced. This criterion rate was held to cut the complex cell orientation response curve at c. 10–20°, the S-shaped effect thereafter being explained as a normal threshold ogive. A similar suggestion seems appropriate here. Later stages of processing could be thought of as scanning the disparity cell population and accepting fairly widespread activity above the criterion as relating to a "real" object or surface. This would mean, in effect, that if each eye's image contained a set of line features whose members fell within areas which interacted binocularly, then the brain would only accept these as deriving from a certain object at a certain depth if the line features had left and right eye orientations within about ± 10° of one another; if this condition was not met then that particular depth hypothesis would be rejected.

As noted in the Introduction, the simple finding that θ differences affect the perception of depth can be accounted for just as well by a point-for-point process as it can by a monocular pattern analysis one. This consideration forces a choice between the two explanations on the basis of how well each deals with the detailed nature of the θ effect, and, we submit, the manner in which the results reported here interlock with expectations based on the physiological findings makes the monocular pattern analysis view the more plausible and parsimonious.

Figure 1 shows the group mean result from the "control" 0/90° condition. This condition was included in the experiment because pilot runs had suggested that the extreme θs did produce a discriminable square set off from the surround even though this square was not seen in depth (see rating 4 above). Now it will be remembered that the stimuli were composed of both black and grey line elements. This meant that all θ conditions, even the extreme ones, possessed some systematic brightness-point disparity because of the unchanging nature of the centres of the lines. It seemed possible that this small resemblance might have proved sufficient for a Julesz point-mechanism to provide an output which conflicted with that from a form-sensitive channel, the two together somehow producing a square but not a depthful one. On the other hand, this square might have been due not to any residual brightness-point disparity but rather to binocular rivalry deriving from non-fusable line elements. A preliminary check on these two possibilities was afforded by the 0/90° pair because they incorporated a square of lines sharing in common an extreme θ difference but without any brightness-point disparity. Predictably, this condition gave rise to "no-depth" but it did nonetheless result in perceptions of a square set off from the surround, showing that the rivalry interpretation could alone have been responsible for the perceived squares in the extreme θ conditions. The work of Kaufman (1964) also supports this interpretation.

The verbatim responses were also reviewed in a qualitative manner. It was evident that when receding depth was seen the central square elements generally appeared larger than those in the surround and, similarly though less noticeably, when the square floated forward its elements were often perceived as smaller. These are clear examples of constancy. Occasionally elements in a forward floating square were also reported as brighter than surrounding ones. The square, as is commonly found in Julesz figures, seemed to have a corporate identity and moved about as a unit when the subject moved his head from side to side. Rather more interestingly, peripheral viewing of the square (by fixating a point somewhat to one side of it) sometimes enhanced the impression of depth. This was reported by five subjects in the 30° condition and by four subjects in the 40° condition, but by no more than two subjects in any other condition. This suggestive result, whose reliability we are currently investigating in a more controlled situation, might be related to the fall-off in acuity that occurs with peripheral viewing, a fall-off which, given our general interpretation of the results, suggests that low peripheral acuity might be due in part to relatively less precise orientational tuning of peripheral striate units.

Since completing this research, the doctoral thesis of Marlowe (1969) has been brought to our attention. His first experiment was very similar to the one described here but he died before publishing his findings in journal form.

References

BARLOW, H. B., BLAKEMORE, C. and PETTIGREW, J. D. (1967). The neural mechanism of binocular depth discrimination. *Journal of Physiology*, **193**, 327–42.

BLAKEMORE C. (1970). The representation of three-dimensional visual space in the cat's striate cortex. *Journal of Physiology*, **209**, 155–78.

BLAKEMORE, C. and CAMPBELL, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, **203**, 237–60.
CAMPBELL, F. W. and KULIKOWSKI, J. J. (1966). Orientational selectivity of the human visual system. *Journal of Physiology*, **187**, 437–45.
CAMPBELL, F. W. and MAFFEI, L. (1970). Electrophysiological evidence for the existence of orientation and size detectors in the human visual system. *Journal of Physiology*, **207**, 635–52.
FRISBY, J. P. (1971). The effects of stimulus orientation on the phi phenomenon. *Vision Research*. In press.
JULESZ, B. (1961). Binocular depth perception and pattern recognition. In CHERRY, C. (Ed.), *Information Theory*. London: Butterworth & Co.
JULESZ, B. (1965). Texture and visual perception. *Scientific American*, **212**, 38–48.
JULESZ, B. (1967). Experiment in Perception. *Readings in Psychology Today*, Communications/Research/Machines, Inc., 212–21.
KAUFMAN, L. (1964). Suppression and fusion in viewing complex stereograms. *American Journal of Psychology*, **77**, 193–205.
KAUFMAN, L. and PITBLADO, C. (1965). Further observations on the nature of effective binocular disparities. *American Journal of Psychology*, **78**, 379–91.
MARLOWE, L. H. (1969). Orientation of contours and binocular depth perception. Ph.D. thesis, Department of Psychology, Brown University.
MITCHELL, D. E. (1969). Qualitative depth localization with diplopic images of dissimilar shape. *Vision Research*, **9**, 991–4.
PETTIGREW, J. D. (1965). Binocular interaction on single units of the striate cortex of the cat. B.Sc. (Med.) thesis, Department of Physiology, University of Sydney.

Received 18 *August* 1971

Quarterly Journal of Experimental Psychology (1971) 23, 373–380

PERCEIVED CURVATURE OF ARCS AND DOT PATTERNS AS A FUNCTION OF CURVATURE, ARC LENGTH, AND INSTRUCTIONS

VEIJO VIRSU
University of Helsinki
AND DANIEL J. WEINTRAUB
University of Michigan

Arcs of circles, with six arc lengths and four radii of curvature, and an equivalent set of figures composed of three dots were used as stimuli. Subjects in Group I imagined the circle from which an arc or dot triplet was taken and indicated the centre of the circle. Group II subjects estimated the location of the point that was equidistant from the middle and ends of an arc, or equidistant from the three dots of a triplet. The results from arcs showed, in Group I, an underestimation of curvature that decreased as a function on the length of the arc. In Group II, however, overestimation of the curvature of most arcs occurred, indicating a strong influence of the difference in the perceptual task on the results. The effect of instructions was similar with the dot figures but, in general, more errors resembling overestimation of curvature occurred with these figures.

Introduction

There are several geometrically equivalent methods for measuring the apparent curvature of a single circular arc. Experimental evidence so far, however, seems to indicate that geometrical equivalence is not sufficient to guarantee perceptual equivalence. Piaget and Vurpillot (1956) and Coren and Festinger (1967) measured the apparent height (the length of the chord) and width (the length of the sagitta, i.e. the perpendicular distance from the midpoint of the arc to the chord joining its ends) of an arc. Their results, based on these linear properties, indicated *overestimation* of the curvature of short arcs. Virsu (1971*b*) compared the apparent curvature of single arcs with the apparent curvature of complete circles and his results suggested *underestimation* of the curvature of short arcs. A similar underestimation seems to result if the apparent continuations of arcs are estimated (Virsu, 1971*a*). Thus, estimation of the curvilinear properties of arcs yields underestimation of curvature. An experiment was designed to clarify the methodological problems involved in the measurement of apparent curvature and to obtain a set of parametric data on perception of curvature.

Method

Stimuli

One set of stimulus figures consisted of 24 single circular arcs, with four different curvatures and six arc lengths. The radii of curvature (radii of the circles determining the arcs)

were 25, 50, 75 and 100 mm. Thus, the curvatures (inverses of the radii of curvature) of the arcs were 1/25, 1/50, 1/75, and 1/100 mm^{-1}, respectively. In terms of the centre angle subtended, the lengths of the arcs were 9, 18, 36, 72, 144 and 288°. The arcs were drawn with black India ink on 21·3 × 27·5 cm sheets of paper. The line thickness was 0·4 mm. One arc was presented on each sheet so that the centre of the circle (not actually shown) from which the arc was taken was approximately in the middle of the sheet. The chord of the arc (not actually drawn) was parallel to the short sides of the sheet.

The other set of stimuli consisted of 24 figures each displaying three dots so that the dots corresponded geometrically to the arcs described above. Two of the dots defined the end points of each arc and one dot corresponded to the middle of each arc. The dots were drawn as small open circles to contribute to the accuracy of later measurements. The outer diameter of each dot was 1·5 mm.

Subjects

The subjects were 32 paid undergraduates from the University of Michigan.

Design

The subjects were divided into two equal groups. One group received instructions to imagine the circle whose circumference was defined by the arc (or dots) and estimate the location of the centre of that circle. The other group was instructed to find the location of the point that was equidistant from the ends and the middle of each arc or equidistant from the three dots.

All subjects estimated both the arc and dot figures. Within each instruction group, eight subjects estimated the dot figures first and then the arc figures, and the eight remaining subjects had the reverse order of figure type. The 24 figures resulting from four curvatures and six arc lengths in each set of figures were randomized independently for each subject.

The figures were presented in four different orientations (arc opening up, down, left, and right), one orientation per subject. The effects of orientation were balanced over the subjects so that two subjects in each instruction group had the same orientation/figure order combination.

Procedure

The stimulus sheets were reproduced by multilith in order to give each subject his own set of 48 figures. The subject was seated at a table covered with a white paper on which the outline of a rectangle was drawn to show the desired location and orientation of the stimulus sheet. The subject placed one sheet at a time inside the rectangle and marked the required point on the sheet. He was encouraged to erase and correct his estimates and draw a small triangle around his final determination. The viewing distance was approximately 40 cm, and the angle between the line of sight and the surface of the sheet was approximately 80°. The instructions, written and illustrated by an appropriate diagram, were continuously available for the subject to consult.

Results

The estimates made by the subjects were measured in mm as two-dimensional deviations from the true centres of the circles from which the arcs were taken. As the orientations of the stimuli were varied, the possible orientation-dependent biases can be evaluated from the individual mean results. There was no statistically significant left–right bias present in the results for dots or arcs, separately or together, in either instruction group or in their combination. There was a clear tendency across all conditions to place the required point too high in the visual field of the subject. The overall individual mean estimates were too high for 23

subjects and too low for nine subjects. As the design was balanced regarding the other effects, a two-tailed binomial test of equiprobability can be used for obtaining a conservative estimate of the statistical significance of the bias and the test yields $P < 0{\cdot}02$. A similar upward bias of point settings has been found also in estimates of the continuations of straight line segments (Weintraub and Krantz, 1971), and therefore, it probably does not indicate any asymmetry in perception of curvature. The effects of presentation order were neither systematic nor statistically significant.

The location of the estimated point on the bisector of the chord of the arc (or on the bisector of the angle formed by the dots) can be interpreted to reflect the perceived radius of curvature of the arc. Therefore, the marks made by the subjects were measured on the bisector as deviations from the geometrically correct centre. If a mark did not lie on the bisector, its perpendicular projection was recorded. The mean results of these measurements are shown in Figure 1. Positive deviations refer to overestimation of the radius of curvature, that is, to underestimation of curvature, and negative deviations refer to underestimation of the radius or overestimation of curvature. Table I presents the standard errors of the means for each radius length.

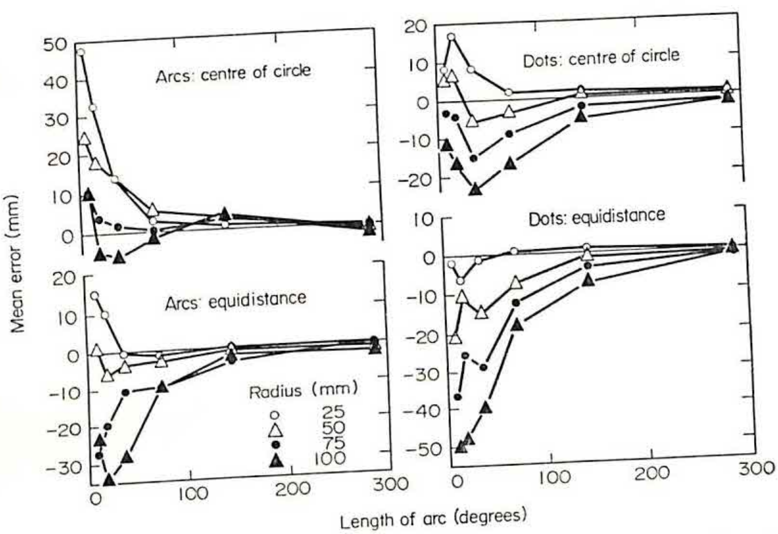

FIGURE 1. The errors in the estimated radius of the arc and dot figures as a function of the arc length (degrees of the centre angle corresponding to the arc) and the radius of curvature.

In general, large and statistically significant errors in estimated radius occurred with short arcs and corresponding dot figures. The largest average error was a nearly 200% overestimation of the radius of the arc with 25 mm radius and 9° length when the centre-of-circle instructions were used.

The errors are clearly affected both by the instructions and by figure type. In arc figures, the underestimation of curvature (overestimation of the radius) typical of the instructions to estimate the centre of the imagined circle tended to change to overestimation of curvature (underestimation of the radius) with instructions

TABLE I

Standard errors of the means (mm)

Radius (mm)	Arc length (degrees) 9	18	36	72	144	288
			Arcs: centre of circle			
25	9·4	8·1	2·9	1·3	0·7	0·2
50	9·7	6·5	3·6	1·8	1·3	0·4
75	8·6	5·6	4·1	2·2	1·8	0·4
100	9·4	4·4	4·8	3·0	1·8	0·7
			Arcs: equidistance			
25	10·5	7·8	3·4	1·3	1·0	0·2
50	9·8	7·3	5·2	2·9	1·8	0·6
75	9·7	8·0	6·8	4·4	3·1	0·4
100	15·2	8·2	6·9	4·6	4·1	0·5
			Dots: centre of circle			
25	6·5	6·7	5·2	2·3	1·3	0·4
50	6·4	7·2	2·5	2·5	2·4	0·6
75	8·2	6·8	3·4	2·1	2·7	1·2
100	8·9	9·1	3·6	3·1	4·4	2·2
			Dots: equidistance			
25	8·8	3·2	3·3	2·2	1·4	0·3
50	9·1	9·4	3·7	2·2	2·8	0·7
75	9·6	8·6	4·6	3·3	3·1	1·0
100	12·4	8·0	7·3	3·7	4·9	1·0

to estimate equidistance. A similar tendency toward more negative radius estimates occurred with dot figures, and in addition, the dot figures yielded more negative errors than the arc figures with either instructions. With equidistance instructions, underestimation of the radius reached nearly 50% for dot figures corresponding to arc figures of long (100 mm) radius and short (9°–18°) arc length. In all the four figure-instructions combinations, however, the overall order of the radii was preserved: the figures with 25 mm radius show the least negative effects.

A rapid decrease in errors occurred in all cases when the length of arc increased. In the arc figures with centre-of-circle instructions, the decrease was approximately exponential. An asymptotical near-zero error was reached with all radii in arc figures with centre-of-circle instructions when the width of the arc was about 1 cm. In terms of visual angle, this value corresponds to approximately 1·4°. The largest arc with which a statistically significant amount of underestimation of curvature was still observed, was that with 72° length and 50 mm radius. In visual angle, the length of the chord (height) of this arc was about 8·5° and the length of the sagitta (width) 1·4°. The mean errors obtained with the 75 and 100 mm radii were not statistically significant.

Discussion

The results of the earlier experiments indicated that different methods produce different outcomes concerning the apparent curvature of single arcs. However, several differences in the measurement arrangements of those experiments could have caused the differing results. In the present experiment, the experimental conditions for the two instruction groups were identical, the only difference being in the perceptual task given to the subjects: one group was instructed to estimate the centre of the circle from which the arc was taken and the other group estimated equidistance from a point. Clear differences in the results of the two groups were still obtained both with arc and dot figures. The subjects responded as if the small arcs or dot patterns looked much more curved in the equidistance estimation than in the centre-of-circle estimation. Evidently, some kind of task-dependent perceptual selection took place even with these simple stimuli so that different deviations of the subjective geometry from the objective geometry occurred with different instructions.

Consistent results have been found with three of the five methods employed so far. Similar amounts of underestimation of curvature resulted from the estimations of the continuations of arcs (Virsu, 1971*a*), comparisons of arcs with circles (Virsu, 1971*b*) and estimations of the centre of curvature.† In these methods, a mental continuation or completion of the arc was required. The important feature of this kind of method is that it is necessary and sufficient to estimate only the curvedness of the arc (that is, the rate of change in the direction of the line), in order to perform the task; the linear aspects of the stimulus are irrelevant for the estimation. The consistency of these judgements suggests that it is possible to perceive the amount of curvature as such, independently of the size features of the arc. The systematic errors are very different if size or distance estimates are required (Coren and Festinger, 1967; Piaget and Vurpillot, 1956). Perhaps there is a specific mechanism for perception of curvature, which is not activated when some primarily positional properties of arcs are estimated. In addition, the clear differences between the arc and dot figures found with the same instructions indicate that the possible mechanism of curvature perception is not fully activated by arcs formed of three dots.

The effects of instructions and figure type are consistent with a mechanism in which perception of curvature is based on the activity of a series of successive orientation detectors, such as those postulated, for example, by Andrews (1967), Blakemore, Carpenter and Georgeson (1970), and Bouma and Andriessen (1970) for perception of angles. The curvature of an arc could be represented by the differences in the orientations of the successive most active orientation detectors so that larger differences indicate greater curvature. Perception of curvature would then be similar to perception of angles except that in the case of arcs the

† For 50 mm radius and 9, 18, 36, 72, 144 and 288° arc lengths, the mean radius errors in the present experiment were 24·9, 18·0, 14·3, 5·2, 2·8 and − 0·6 mm. The corresponding averaged values from Virsu (1971*b*) are 23·2, 17·1, 10·1, 3·2, 2·2 and 1·5 mm (the curvatures of arcs with 50 mm radius were matched with the curvatures of circles). The radius errors derived in Virsu (1971*a*) were 9·0 mm for 30° arc length and 0·9 mm for 90° arc length (the intersection of the continuations of two arcs with 50 mm radius was estimated).

higher neural mechanisms would integrate orientation information over a *large number of different* orientations while for angles the integration would concern *two orientations only.*

It can be assumed that there are many orientation detectors tuned for different orientations at any location of the visual field and many parallel tuned detectors for adjacent positions. Both the position and orientation of lines could be indicated by the orientation detectors. Then, as long as the retinal locus of a line segment remains the same, the position information conveyed by the detector outputs is the same but the orientation of the most activated detector can change. Also the position information can change without a change in the orientation information. The detector outputs for position and orientation must be independent in some sense because we can perceive, for instance, that a line segment rotates in a fixed position in the visual field or that a line segment moves without changing its orientation.

It was assumed by Blakemore *et al.* (1970) that lateral inhibition can take place in the orientation domain between the orientation detectors so that it changes the orientation of the most activated detectors. In the system above, any modification of the orientation information caused by inhibition, adaptation or summation, which is restricted to the orientation detectors, would be similar to a veridical change of orientation and would not affect the position information conveyed by the detector outputs. Therefore, an error in perceived orientation can occur without an accompanying error in perceived position. Similarly, an inhibitory or other shift in the position domain does not alter the orientation information.

The task-dependent perceptual selection could mean the selection of the output that is relevant to the responses required and inhibition of the irrelevant outputs. In the present experiment, the relevant output for determining the equidistance responses was the position output, and for the centre-of-circle responses, in which a mental completion of the arc was required, the relevant output was the orientation output. The fact that systematic errors were observed in each case indicates that error-producing processes take place both in the orientation and position domain. The inconsistency of the perceptual outcomes is not surprising for it would be most unlikely that an error-producing shift in the orientation of the most activated detectors would be accompanied by a shift in the position domain such that the perceptual consequences of these shifts would be geometrically equivalent. Inconsistencies of perceived relative position and orientation can be expected in all cases in which the stimulus evokes a nonveridical position or orientation shift and the experimental arrangements are suitable for revealing the inconsistency. Morinaga's paradox (based on the Müller-Lyer figure, see Oyama, 1960) and the apparent incompatibilities of the visual geometry discussed by Rubin (1950) and Virsu (1971*b*, see Figure 3) are demonstrations of this rule.

The difference between the results from arc and dot figures can be explained, for it is likely that the dot figures did not form fully adequate stimuli for the postulated orientation detectors. The receptive fields of the orientation detectors may be quite small in the central fovea: Andrews (1967) estimated that the smallest fields cover only 9 min of arc. The field sizes vary considerably, however. Therefore, assuming that one dot at each end of the receptive field is an effective

stimulus for a detector, some similarity in the results for arcs and dots could be expected. Only with the smallest dot figures, however, could the stimulation of the detectors have been strong enough to produce orientation interactions (in the smallest figure the dot intervals were about 16 min). This feature can be seen in the data: the curves for dots with centre-of-circle instructions descend first like the curves for arcs with centre-of-circle instructions but then they ascend like the curves for dots with equidistance instructions.

Three possible explanations for underestimation of curvature may be mentioned. First, in the previous study (Virsu, 1971*b*) the interpretation was considered that tendencies to rectilinear eye movements could cause underestimation of curvature. Even though this explanation is consistent with the error functions, the explanation is not very satisfactory because it is neither economical nor directly testable. The explanation requires two mechanisms: one for the initial perception of curvature and another for the errors. A second explanation is that inherent differences in the tuning characteristics of the orientation detectors for oblique and horizontal or vertical orientations cause underestimation of the curvature of horizontal and vertical arcs; such tuning differences have been postulated by Andrews (1967) and Bouma and Andriessen (1970) for explaining perceptual tilt of straight line segments. Yet, this obvious possibility is not a sufficient explanation, for the underestimation of curvature is only slightly smaller with oblique arcs than with vertical arcs (Virsu, 1971*b*). The third possibility is that underestimation of curvature represents an initial stage of the Gibson effect, in which a curved line becomes perceptually straightened with prolonged inspection (Gibson, 1933). If this were the case, some kind of neural adaptation might explain the results. The difficulty of this explanation is that the Gibson effect seems to occur typically with very long radii and with arcs that are large in terms of visual angle. In the present experiment, however, the underestimation of curvature occurred only with the shortest arcs and radii: the errors for the 75 and 100 mm radius lengths were not statistically significant. On the other hand, the inspection times might not have been long enough for the development of the Gibson effect.

Supported by Kamarineuvos H. Rosenberg's Travel Grant from the University of Helsinki to V. V. and United States Public Health Service Research Scientist Development Award (K2-MH-35, 253) to D. J. W. Research Funds were provided by National Science Foundation Grant GB 8181 and by the University of Michigan's Biomedical Sciences Institutional Support Grant PHS No. FR 07050-03. We thank Lillian Tong for serving as Experimenter.

References

Andrews, D. P. (1967). Perception of contour orientation in the central fovea. Part II. Spatial integration. *Vision Research*, **7**, 999–1013.

Blakemore, C., Carpenter, R. H. S. and Georgeson, M. A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, **228**, 37–9.

Bouma, H. and Andriessen, J. J. (1970). Induced changes in the perceived orientation of line segments. *Vision Research*, **10**, 333–49.

Coren, S. and Festinger, L. (1967). An alternative view of the "Gibson normalization effect". *Perception and Psychophysics*, **2**, 621–6.

GIBSON, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *Journal of Experimental Psychology*, **16**, 1–31.

OYAMA, T. (1960). Japanese studies on the so-called geometrical optical illusions. *Psychologia*, **3**, 7–20.

PIAGET, J. and VURPILLOT, E. (1956). La surestimation de la courbure des arcs de cercle. *Archives de Psychologie*, **35**, 215–32.

RUBIN, E. (1950). Visual figures apparently incompatible with geometry. *Acta Psychologica*, **7**, 365–87.

VIRSU, V. (1971*a*). Tendencies to eye movement, and misperception of curvature, direction and length. *Perception and Psychophysics*, **9**, 65–72.

VIRSU, V. (1971*b*). Underestimation of curvature and task dependence in visual perception of form. *Perception and Psychophysics*, **9**, 339–42.

WEINTRAUB, D. J. and KRANTZ, D. H. (1971). The Poggendorff illusion: Amputations, rotations, and other perturbations. *Perception and Psychophysics*, **10**, 257–64.

Received 30 *April* 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 381–385

SYMMETRY, GESTALT AND INFORMATION THEORY

J. B. DEREGOWSKI
University of Aberdeen

Two kinds of patterns were made up out of crosses. The first kind was symmetrical and the second contained a sub-pattern repeated in the same orientation. Subjects were requested to reproduce the stimuli after viewing them for 2 sec. It was found that there was no difference between the two kinds of stimuli when symmetry about the horizontal axis was involved but that when symmetry about the vertical axis occurred the symmetrical stimuli were reproduced more accurately than the repeated stimuli. Introduction of contrasting colours on the two portions of each pattern decreased the frequency of correct responses, whilst stressing the grid line subdividing the stimuli into two portions containing the same information had no significant effect.

Introduction

Both major attempts at quantification of the Gestalt laws, that of Hochberg and McAlister (1953) and that of Attneave (1954), agree that symmetry, which Koffka considered to be one of the important determinants of figural "goodness", plays an important role in perception. Attneave defines symmetry as a form of redundancy (1954, p. 166) and in a subsequent study (1955) he shows that symmetrical patterns are better remembered than non-symmetrical patterns containing the same number of cells. He notes that information measurement offers an explanation of the phenomenon but that this explanation is imperfect because the data obtained show that there is an increase in number of errors with an increase in number of cells when the information (measured in bits) is held constant.

Examination of Attneave's graphs suggests that this increase is not linear, there being a smaller rise in errors on increasing the pattern from 20 to 35 cells than on increasing it from 10 to 20 cells. The difference is probably attributable to the different type of symmetry these patterns involve. The 20 cell pattern is symmetrical about the vertical axis whereas the 35 cell pattern is symmetrical about both the vertical axis and the horizontal axis. Paraskevopoulos (1968) observed that the latter type is much more readily recalled than the former. This suggests that symmetry may involve some factors which traditional information measurement fails to account for.

This impression is supported by the experimental evidence of Fitts *et al.* (1956) who showed that mirror-image figures were recognized more rapidly than "double contour asymmetrical figures". The latter had the same information value as the symmetrical (mirror-image) figures and consisted of a repetition of the same pattern in two halves of a matrix instead of rotation out of the plane of the paper which occurs in symmetrical figures. Further, Fitts *et al.* and Paraskevopoulos

found that the orientation of the axis of symmetry affected the response. Thus, figures symmetrical about the vertical axis were recognized more quickly than the same figures so rotated that they were symmetrical about the horizontal axis (Fitts *et al.*, 1956); and similarly children aged between 7 and 11 years of age made significantly more errors in reproducing patterns symmetrical about the horizontal axis than they did in reproducing patterns symmetrical about the vertical axis (Paraskevopoulos, 1968). There are therefore differences in responses to matrices having the same number of cells and containing the same number of bits.

It is possible to interpret the superiority of symmetrical patterns over repeated patterns by assuming that the former are more easily divisible into identical subsets; moreover, such a division is easier in patterns symmetrical about the vertical axis than it is in patterns symmetrical about the horizontal axis. Hence one would postulate that any modification of stimuli which stresses the existence of subsets would facilitate perception and recall of these stimuli. On the other hand, from a "Gestalt" point of view introduction of such a modification to stimuli should tend to destroy their "goodness" and hinder these processes.

The above considerations lead to the following hypotheses:

(1) a stimulus symmetrical about the vertical axis will be found less difficult to recall than the corresponding repeated stimulus;

(2) it will also be less difficult to recall than a stimulus symmetrical about the horizontal axis;

(3) contrary to Gestalt theory, an increase in conspicuousness of the two elements from which the stimuli are constructed either by varying the colouring of the elements or by introducing a distinct dividing line will lead to more correct responses.

In addition since it has also been observed that responsiveness to symmetry in matrices of this type varies with age (Paraskevopoulos, 1968), younger children treating all stimuli whether symmetrical or not in a like manner, a further hypothesis was put forward:

(4) the frequency of correct responses will increase with the age of the subjects.

Method

Subjects

The subjects were children from a primary school in a Scottish town. They were derived from three age groups (7–8; 8–9; 9–10), each group yielding 24 subjects and covering a different school class. (In one case it was found necessary to sample one subject from another class of the same grade and to extend the age limit by eight days to obtain the full sample.) Both girls and boys were used.

Apparatus

Six 4 × 4 matrices were prepared by allocating to each cell of the left-hand half of each matrix an X or by leaving it blank, the two outcomes being equiprobable. These "familial" matrices were then used in devising all the five forms of stimuli used. For symmetrical stimuli the pattern arrived at was reproduced symmetrically in the right-hand half and the matrix either left in this orientation (yielding a stimulus symmetrical about the vertical

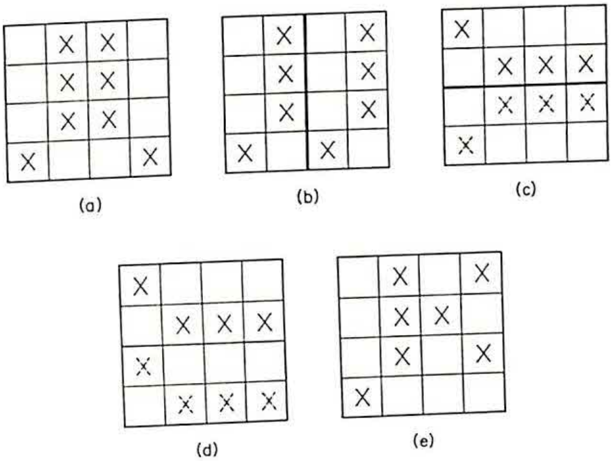

FIGURE 1. Five stimuli of the same family but belonging to different sets: (a) symmetrical about vertical axis, cohesive and in one colour; (b) vertically repeated, divided and in one colour; (c) symmetrical about horizontal axis, divided and in two colours; (d) horizontally repeated, cohesive and in two colours; (e) random, cohesive and in one colour.

axis) or rotated through 90° clockwise (yielding a stimulus symmetrical about the horizontal axis). Repeated pattern stimuli were analogously obtained by repeating the patterns and either leaving the pattern unrotated (vertical repetition) or rotating it through 90° (horizontal repetition). Random patterns were obtained by filling the cells of the right-hand half of the matrix using the same method as that used for the left-hand half. Five families of patterns consisting of vertically symmetrical, vertically repeated, horizontally symmetrical, horizontally repeated, and random were thus created (see Fig. 1). All these families were used in construction of each of four sets of stimuli made up from a 2 × 2 factorial design. The two factors were (i) colour (both halves of a figure being either of the same colour or differing in colour); (ii) the presence or absence of a heavy line separating the two halves of a pattern (yielding "divided" and "cohesive" stimuli respectively). Of the uniformly coloured stimuli three families were black and the other three red. In three of the two-colour families the "familial" patterns were red and the derived patterns black; in three others the colours were interchanged. Two training patterns, both random, one having red X's, the other both red and black and a thickened dividing line, were also prepared.

All the stimulus drawings were 6 cm squares. A "Cambridge" tachistoscope was used. In this apparatus stimuli are presented at about 40 cm from the subject.

Procedure

The subjects were tested individually. Each subject was given a seven-page booklet and told that a picture would be shown to him and he would have to copy it in the booklet. The first introductory stimulus (the red random pattern) was then exposed for a short time on the table in front of the subject. When the subject had completed his response, the second introductory stimulus (two colour, divided random pattern) was presented to him in a tachistoscope. After the two introductory stimuli a set of five test stimuli were presented. Two second exposures were used throughout, and each response was recorded on a fresh page in the booklet.

Each subject was allocated to one of four stimulus conditions: one colour cohesive; two colour cohesive; one colour divided and two colour divided.

Within each set he was asked to respond to a series of five experimental stimuli. The sequence within the experimental series was arranged as follows: each stimulus in a series was derived from a different family. The random stimulus was always placed third and the pair of stimuli preceding it and the pair which followed consisted of one "Horizontal"

(either symmetrical or repeated) stimulus and one "Vertical" (either symmetrical or repeated) stimulus. The order within the pairs alternated from sequence to sequence. Each pair contained one symmetrical and one repeated stimulus. Further, not less than two and not more than three stimuli within each series came from either of the two colours (in the case of the single colour stimuli) or from either arrangement of two colours in the case of the two colour stimuli.

Within these restrictions the stimuli were allocated at random and of six series of stimuli thus obtained each was used once with each subgroup of six subjects.

The subjects were tested in their school.

Results and Discussion

The number of wholly correct responses (for each sub-group of six subjects) was cast into a [3 (age group) × 2 (colour conditions) × 2 (cohesiveness conditions) × 5 (replications within a subject)] table, and subjected to analysis of variance using a Split plot (SPF pru.q) design with highest order interactions as error terms (Kirk, 1968). None of the lower order interaction terms was found to be significant. Two of the main effects were significant: type of stimulus ($F = 11 \cdot 8$; $df = 4, 8$, $P < 0 \cdot 005$), and the effect of uniformity and non-uniformity of colour ($F = 28 \cdot 3$, $df = 1, 2$, $P < 0 \cdot 05$). The age effect approached significance ($F = 13 \cdot 7$, $df = 2, 2$, $P < 0 \cdot 10$). There was a lower number of correct responses in the youngest group.

As the differences between the three types of stimulus were highly significant, three further comparisons between pairs of stimuli were carried out. The correct responses to the random stimuli were compared with those made to the stimulus from which they differed least, the vertical repeated stimulus, and were found to be significantly less frequent ($t = 2 \cdot 3$, $df = 11$, $P < 0 \cdot 025$). A comparison was also made within each pair of similarly orientated non-random stimuli. It was found that there was a significant difference between the vertically repeated and the vertically symmetrical stimuli ($t = 5 \cdot 4$, $df = 11$, $P < 0 \cdot 0005$) the former being reproduced correctly less often. There was no significant difference between the corresponding horizontal stimuli ($t = 1 \cdot 08$, $df = 11$, $P < 0 \cdot 20$).

TABLE I

Average number of Correct responses

Colour	Vertical stimuli		Horizontal stimuli		Random stimuli
	Symmetrical	Repeated	Symmetrical	Repeated	
One Colour	0·53	0·19	0·31	0·28	0·03
Two Colours	0·39	0·11	0·25	0·17	0·03

Table I presents mean frequencies of correct responses obtained tabulated with respect to the two significant variables.

The results support the first hypothesis, in that the stimulus symmetrical about the vertical axis was found to be less difficult than its repeated counterpart. This

suggests that the symmetry of a figure facilitates responding significantly more than a mere repetition does and hence that the information theory measure does not provide a sufficient tool for measurement of figural characteristics.

This limitation of information theory cannot, however, be said to apply to the horizontal figures as well, for in this case the postulated hypothesis obtains no support, as there is no significant difference between the two stimuli, although there is a slight trend in the same direction as in the case of the "vertical" stimuli.

This absence of a significant difference is contrary to the results reported by Fitts *et al.* (1956). Fitts, however, used a sample of adults, a different procedure and a different response measure. Paraskevopoulos' findings that the extent to which symmetry is used increases with age suggest that the lower age of subjects used in the present experiment may be chiefly responsible for the non-significance of the observed difference.

Of the two methods introduced to break up figural gestalt into two portions containing identical information load, the division of the matrix produced no significant effect, whilst the use of two colours led to fewer correct responses. This is surprising, for it appears that this difference, which was intended to bring our identity of patterns in the two units, has just the opposite effect. The result therefore favours the Gestalt interpretation in terms of figural goodness of patterns symmetrical about the vertical axis.

Since the "number of colours $\times$ type of stimulus" interaction was non-significant, the results suggest that the increase of the number of errors with introduction of two colours did not differ between the symmetrical and the repeated stimuli. Since a significantly smaller number of errors was made in response to repeated stimuli than in response to random stimuli, it further suggests that subjects must make use of the repeated sets, but that they use them less efficiently when differences in colour occur.

As the Gestalt goodness which applies to the symmetrical figures cannot be easily invoked in the case of the repeated stimuli one must conclude, rather unsatisfactorily, that neither the Gestalt nor the information theory approach offers a satisfying explanation of this aspect of the data.

References

ATTNEAVE, F. (1954). Some informational aspects of visual perception. *Psychological Review*, **61**, 183–93.

ATTNEAVE, F. (1955). Symmetry, information and memory for patterns. *American Journal of Psychology*, **68**, 209–22.

FITTS, P. M., WEINSTEIN, M., RAPPAPORT, M., ANDERSON, N. and LEONARD, J. A. (1956). Stimulus correlates of visual pattern recognition: a probability approach. *Journal of Experimental Psychology*, **51**, 1–11.

HOCHBERG, J. and MCALISTER, E. (1953). A quantitative approach to figural "goodness". *Journal of Experimental Psychology*, **46**, 361–4.

KIRK, R. E. (1968). *Experimental Design: Procedures for Behavioural Sciences.* Belmont: Brooks/Cole.

PARASKEVOPOULOS, I. (1968). Symmetry recall and preference in relation to chronological age. *Journal of Experimental Child Psychology*, **53**, 237–43.

Received 18 *August* 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 386–392

DICHOTIC BACKWARD MASKING OF COMPLEX SOUNDS

C. J. DARWIN†

Haskins Laboratories, New Haven

In the first experiment subjects identified a consonant–vowel syllable presented dichotically with a known contralateral masking sound at a stimulus onset asynchrony of $\pm$ 60 msec. When the mask followed the target syllable, perception of place of articulation of the consonant was impaired more when the mask was a different consonant–vowel syllable than when it was either a steady-state vowel or a non-speech timbre. Perception was disturbed less when the mask preceded the target, and the amount of disruption was independent of which mask was used. Greater backward than forward masking was also found in the second experiment for the identification of complex sounds which differed in an initial change in pitch. These experiments suggest that the extraction of complex auditory features from a target can be disrupted by the subsequent contralateral presentation of a sound sharing certain features with the target.

Introduction

The traditional task in experiments on the temporal course of auditory masking has been the *detection* of a target presented in close temporal proximity to a mask. This paradigm has shown only small effects when target and mask are presented to opposite ears (dichotically). Moreover, these effects have been found only over very brief stimulus-mask intervals. Elliott (1962), for example, found virtually no forward masking of a brief tone by contralateral white noise, and only slight backward masking extending out to an inter-stimulus interval of about 15 msec.

Recently Studdert-Kennedy, Shankweiler and Schulman (1970) have reported an experiment requiring *identification* of two stop–consonant syllables presented dichotically with a temporal offset between them. They found that for offsets between about 15 and 120 msec the lagging syllable was reported more accurately than the leading syllable. Their result has since been confirmed both in the original paradigm (Berlin *et al.*, 1970; Lowe *et al.*, 1970) and in a slightly different paradigm in which only one sound has to be reported on a single trial (Kirstein, 1970, 1971). No advantage for the lagging over the leading sound, however, was found in binaural presentation (with both syllables coming to both ears) even when the duration of the vowel portion of each syllable was drastically curtailed to eliminate temporal overlap between the two sounds (Porter, 1971). Such curtailing did not influence the dichotic effect.

In terms of masking, these experiments have shown that under dichotic presentation stop-vowel syllables are more effective as mutual backward than as mutual forward maskers, whereas under binaural presentation, provided they do not temporally overlap, any masking that occurs is symmetrical.

† Now at University of Sussex.

In the visual modality, dichoptic masking is essentially a contour interaction (Schiller, 1965; Kahneman, 1968), which is asymmetrical so that backward masking is greater than forward. This asymmetry supports theories which emphasize the interruption of perceptual processes by the mask, rather than a temporal summation of mask and target (Kahneman, 1968; Spencer and Shuntich, 1970). A similar explanation seems appropriate for the auditory case (Studdert-Kennedy, Shankweiler and Schulman, 1970), although for stop–vowel syllables the effect is confined to dichotic presentation, whereas in vision monoptic contour interactions can be obtained (Schiller, 1965).

The present study pursues the analogy between dichoptic and dichotic masking. In the auditory experiments reviewed above it is not clear whether the more effective backward than forward masking is confined to a particular type of mask, since only syllables have been used to mask syllables. The first experiment therefore examines the relative extent of forward and backward masking for a number of different masks on a stop-vowel target set.

Experiment I

The masks used in this experiment were chosen to have certain properties in common with the target set. Three were speech and the fourth a non-speech timbre. The three speech sounds were (a) a steady-state vowel different from that used in the target syllables, (b) the same vowel as used in the target syllables and (c) a stop-vowel syllable with the same vowel as the targets but a different stop consonant.

Method

The targets used in this experiment were the four stop-vowel syllables /*bɛ*, *dɛ*, *pɛ*, *tɛ*/. These four consonants give two values each on the traditional phonetic dimensions of place of articulation and voicing. The four masks were /*gɛ*, *ɛ*, /*ɔ* and a non-speech steady-state timbre, which had three formants at 894, 2910 and 3698 Hz respectively. The two steady-state vowel masks and the five syllables were all highly intelligible. All the sounds were synthesized with three formants on the Haskins parallel formant synthesizer. Each sound lasted 100 msec and all the sounds had the same intonation contour and were equated for peak amplitude on a VU meter. On each trial of the experiment a subject heard one of the target sounds in one ear and one of the masks in the other. He always knew which mask would occur since this was held constant over a block of 48 trials and was played to him before each block, but he did not know into which ear the target would come. His task was simply to identify which of the four targets had been presented; he did not have to say into which ear it had come. The sounds on the two ears were always temporally offset by 60 msec. Whether the target or the mask led was randomly determined with the restriction that within each block of 48 trials each target item led six times and lagged six times. 16 subjects each took eight blocks of 48 trials in a Latin square design which counter-balanced the order in which the four masks were heard. The subjects were given a binaural demonstration of the set of target items before taking the dichotic test. Before each block the mask for that block was played three times binaurally.

Results

Three different scoring methods were used: (a) the response had to be entirely correct (both place of articulation and voicing); (b) only place of articulation need be correct; and (c) only voicing need be correct. The results according to these

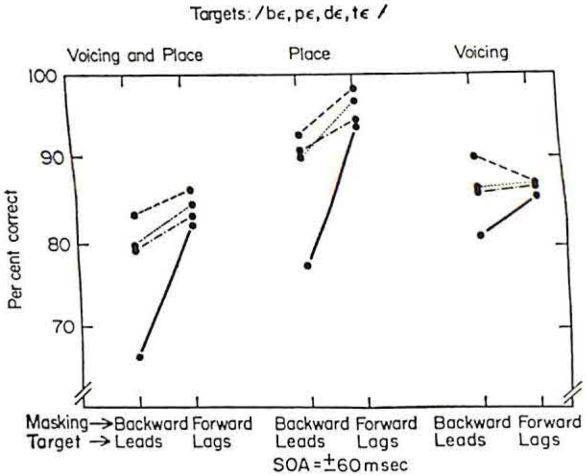

FIGURE 1. Mean percents correct for stop-vowel syllables dichotically opposed by a mask at ± 60 msec offset. The three columns refer to three different scoring criteria
Mask: gɛ — ɛ · · · ·; ɔ — · —. Non-speech — —.

three methods are shown in Figure 1. The slope of each line indicates the difference between the target leading and target lagging conditions for the various masks. A line with a positive slope indicates that the target is better perceived when it lags the mask than when it leads it.

In the scores where both place of articulation and voicing had to be correct, analysis of variance showed a significant interaction of the lead/lag factor with mask (F (3, 105) = 3·98, $P < 0{\cdot}01$). However, since an analysis of variance on the results for place of articulation and voicing separately showed a significant difference between these two features for the interaction of lead/lag and mask (F (3, 45) = 3·16, $P < 0{\cdot}05$) as well as a significant interaction between the feature analysed and lead/lag condition (F (1, 15) = 23·8, $P < 0{\cdot}001$), the results will now be described separately for these two features.

For place of articulation, as with both features combined, there was a significant interaction of mask with whether the target led or lagged the mask (F (3, 45) = 12·5, $P < 0{\cdot}001$). However, as is clear from the figure, this interaction is mainly due to the case when the target leads the mask (i.e. to the backward masking case). This was confirmed in analysis of variance which showed a highly significant effect of mask on a preceding target (F (3, 45) = 18·6, $P < 0{\cdot}001$), but only slight variation when the target follows the mask (F (3, 45) = 2·75, $0{\cdot}1 > P > 0{\cdot}05$). Thus for perception of place of articulation the amount of forward masking is virtually independent of the mask, but the amount of backward masking is much greater when the mask is another stop–vowel syllable than when it is one of the other masks ($P < 0{\cdot}001$). However, the three steady-state masks do show significantly greater backward than forward masking ($P < 0{\cdot}001$) although the amount of backward masking is very much less than for /*gɛ*/.

For the extraction of voicing, however, there was no overall advantage for the lagging over the leading target ($F < 1{\cdot}0$) and only a slight interaction of lead/lag condition with mask (F (3, 45) = 2·46, $0{\cdot}1 > P > 0{\cdot}05$). Thus the perception of voicing shows no more backward than forward masking for the masks used here.

In summary this experiment shows that for stop–vowel syllables dichotically opposed by a mask at temporal offsets of $\pm$ 60 msec: (1) forward masking is roughly constant for the four masks used, for both place of articulation and voicing; (2) backward masking is greater than forward with place of articulation although not with voicing for all the masks; but (3) this difference is considerably greater when the mask is another stop–vowel syllable than when it is either the same vowel, a different vowel or a non-speech timbre; (4) these last three masks do not differ significantly in any condition.

Discussion

The amount of backward masking, at least for the perception of place of articulation, is clearly dependent on the mask used. Dichotic masking is thus a potentially useful tool for describing features in auditory perception. The sharp discontinuity between the effects of the three steady-state masks and the /*gε*/ mask argues against any general continuum of similarity being important, for if it were we might have expected the /ε/ mask to have been closer in its effect to the /*gε*/ mask. Rather, we are led to suppose that the /*gε*/ mask contains specific features which are particularly effective at interrupting the perception of the preceding target. This interpretation is strengthened by the absence of any mask specificity in the forward masking case, although this may be at least partly due to the very high performance leaving little room for improvement.

Two more points require discussion: the slight though consistently greater effect of backward over forward masking for the three steady-state masks, and the absence of any differences either between masks or between the forward and backward conditions for the perception of voicing. The first point may be attributable to some unspecific auditory effect or perhaps may not even be specific to the auditory modality. A kick on the shins may be an effective backward mask to this extent. The effect is quite small and will probably be difficult to investigate. The absence of any interesting effects in the perception of voicing may reflect the very different acoustic cues underlying the perception of voicing and of place of articulation. For voicing, at least in this experiment, the detection of some aspiration at the beginning of the stimulus would give sufficient information, whereas for place of articulation detailed knowledge of the slope of rapid formant transitions is required. The extraction of this latter information may be particularly sensitive to disruption.

This experiment alone cannot decide whether extraction of those acoustic parameters on which the decision concerning place of articulation is based is being disturbed, or rather whether it is some specifically linguistic process such as the relation of these acoustic features to a phonemic framework. To distinguish between these two hypotheses the next experiment looks at backward masking for stimuli which, like stop–vowel syllables, are distinguished by a rapidly changing initial section, but which are not perceived as falling into different phonemic categories.

Experiment II

This experiment uses a paradigm introduced by Kirstein (1970). No *a priori* distinction is made between target and mask, both being drawn from the same

stimulus set. The subject attends to one ear and is asked to recall the stimulus presented there.

Method

Three different sounds were used. They differed in their fundamental frequency contours which are illustrated in Fig. 2. These pitch contours were carried on the steady vowel /ɛ/. Dichotic pairs were made up using the Haskins parallel formant synthesizer and a special computer program (Mattingly, 1968) which ensured perfect timing of the signals on the tape's two tracks. On each trial a subject heard two pitch contours, one in either ear. They were either simultaneous or offset by ± 25 msec. Subjects attended to one ear for a block of trials and were instructed to identify only the sound that was presented to that ear. They were given training in identifying the three sounds with the first three digits. Half the pairs of sounds they heard were simultaneous and half were temporally offset. 12 right-handed subjects took the experiment in a procedure which counterbalanced ears and attention.

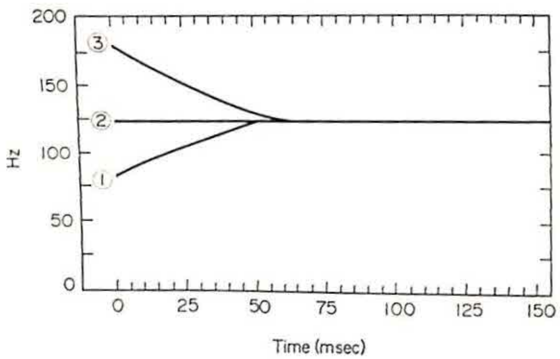

FIGURE 2. Pitch contours of the three sounds used in Experiment II.

Results

The results are tabulated (Table I) in terms of the asynchrony of the *reported* stimulus. Thus, if the subject was presented with stimulus 1 to his left ear 25 msec ahead of stimulus 2 to his right and, though asked to report the left ear, in fact wrote "2", a correct response would be entered in the cell where the "right-ear" column under "lagging" and the row for "attend left" intersect. There was a clear advantage for the lagging over the leading condition irrespective of ear or attention condition (12 subjects for, none against). Subjects were

TABLE I

Mean per cents correct in Experiment II according to asynchrony of the reported stimulus

	Asynchrony of reported stimulus					
	Simultaneous		Leading		Lagging	
	Left	Right	Left	Right	Left	Right
Attend left	40·4	37·1	35·2	33·0	48·6	43·6
Attend right	36·6	40·4	29·5	37·8	40·9	48·6
Total	38·5	38·7	32·4	35·4	44·8	46·1

generally poor at selecting the requested ear though there was some indication that selective attention was easier for the staggered pairs than for the simultaneous ($P < 0.1$). There was no difference between the ears in either the simultaneous or the staggered condition ($P > 0.1$).

Discussion

As in the first experiment we find here greater backward than forward dichotic masking for sounds which are distinguished by a rapidly changing initial portion. In the first experiment, rapidly changing formant transitions cued the place-of-articulation distinction in stops whereas in this experiment the sounds have been distinguished by changes in fundamental frequency which did not cue a phonemic distinction. Parsimony suggests that explanations for these effects should be sought at a purely auditory level of analysis rather than supposing that separate explanations are required for both speech and non-speech sounds.

Two subsidiary results of the second experiment call for comment: first, the slightly more efficient selective attention under staggered than under simultaneous conditions bears out a suggestion by Treisman and Riley (1969) to that effect. Second, the absence of any ear difference here contrasts in an interesting way with Haggard and Parkinson's (1971) finding that when rapid pitch changes similar to the ones used here cue the voiced-voiceless distinction in stops (Haggard, Ambler and Callow, 1970) there is an advantage for the right ear under dichotic presentation. The right ear advantage is thus determined by the use to which acoustic information is put rather than to the presence of particular acoustic features (Darwin 1971; Haggard and Parkinson, 1971).

Both experiments reported here have used dichotic presentation. As mentioned in the Introduction, the superiority of backward over forward masking for stop–vowel syllables by similar syllables is not found for monaural presentation (Porter, 1971). A related finding is that by Massaro (1970), who finds slightly larger dichotic than monotic backward masking for the identification of a pure tone followed by a longer masking tone. This generally greater backward masking under dichotic than under monotic conditions may reflect a segmentation problem faced by the auditory system. If the arrival of a particular auditory feature always interrupted the processing of any similar feature which had arrived just previously, the range of sounds which the auditory system could analyse successfully would be severely restricted. On the other hand, there may be good reasons why interruption should occur when the second sound forms part of a different perceptual unit from the first. Misplaced interruptions would be restricted to a minimum, at least in natural situations, if sounds from the same source were treated without interruption. Spatial location could provide a very reliable criterion for determining whether temporally distinct sounds originated from a common source. Spatial location has the added advantage that at least its directional aspect has neurophysiological correlates at a very peripheral level of the auditory system. This information is thus potentially available for guiding the sequential analysis of the auditory input at higher levels. This may not be the only criterion, and indeed Massaro's experiments with pure tones show appreciable monotic backward masking. But different processes may be operating for simple and complex

stimuli, since Massaro also finds backward masking for pure tones to be relatively independent of the similarity of test and masking tones.

This work was partially supported by a grant to Haskins Laboratories from the National Institute of Child Health and Development while the author was supported by a Fellowship from the Commonwealth Fund of New York.

References

BERLIN, C. I., WILLETT, M. E., THOMPSON, C., CULLEN, J. K. and LOWE, S. S. (1970). Voiceless versus voiced CV perception in dichotic and monotic listening. *Journal of the Acoustical Society of America*, **47**, 75(A).

DARWIN, C. J. (1971). Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology*, **23**, 46–62.

ELLIOTT, L. L. (1962). Backward masking: monotic and dichotic. *Journal of the Acoustical Society of America*, **34**, 1108–15.

HAGGARD, M. P. and PARKINSON, A. M. (1971). Stimulus and task factors as determinants of ear advantages. *Quarterly Journal of Experimental Psychology*, **23**, 168–171.

HAGGARD, M. P., AMBLER, S. and CALLOW, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, **47**, 613–7.

KAHNEMAN, D. (1968). Method, findings and theory in studies of visual masking. *Psychological Bulletin*, **70**, 404–25.

KOLERS, P. and ROSNER, B. S. (1960). On visual masking (metacontrast). *American Journal of Psychology*, **73**, 2–21.

KIRSTEIN, E. F. (1970). Selective listening for temporally staggered dichotic CV syllables. *Journal of the Acoustical Society of America*, **48**, 95(A).

KIRSTEIN, E. F. (1971). Temporal factors in perception of dichotically presented stop consonants and vowels. Ph.D. dissertation, Univ. of Connecticut.

LOWE, S. S., CULLEN, J. K., THOMPSON, C., BERLIN, C. I., KIRKPATRICK, L. L., and RYAN, J. T. (1970). Dichotic and monotic simultaneous and time-staggered speech. *Journal of the Acoustical Society of America*, **47**, 76(A).

MASSARO, D. M. (1970). Preperceptual auditory images. *Journal of Experimental Psychology*, **85**, 411–17.

MATTINGLY, I. G. (1968). Experimental methods for speech synthesis by rule. *I.E.E.E. Transactions*, AU–16, 198–202.

PORTER, R. J. (1971). The effect of temporal overlap on the perception of dichotically and monotically presented CV syllables. Paper at 81st Meeting of the Acoustical Society of America, Washington, D.C.

SCHILLER, P. H. (1965). Monoptic and dichoptic visual masking by patterns and flashes. *Journal of Experimental Psychology*, **69**, 193–9.

SPENCER, T. J. and SHUNTICH, R. (1970). Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology*, **85**, 198–203.

STUDDERT-KENNEDY, M., SHANKWEILER, D. P. and SCHULMAN, S. (1970). Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. *Journal of the Acoustical Society of America*, **48**, 599–602.

TREISMAN, A. M. and RILEY, J. G. A. (1969). Is selective attention selective perception or selective response? A further test. *Journal of Experimental Psychology*, **79**, 27–34.

Received 3 August 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 393–398

IMAGERY AND COMPATIBILITY OF PAIRING IN THE FREE RECALL OF ADJECTIVE-NOUN PAIRS

P. E. MORRIS AND R. L. REID
Department of Psychology, University of Exeter

Can the image-evoking value (*I*) of single words be used to predict the recall of grammatical units in which they occur? Thirty-two subjects free-recalled a list of 24 adjective-noun pairs varying in the *I* of the adjective and the noun, and the compatibility (*C*) of the pairing (i.e. whether or not the pair was immediately meaningful). Analysis showed all three variables to affect recall. High *I* adjectives facilitated recall irrespective of the *I* of the noun or the *C* of the pairing. Noun *I* and *C* facilitated recall only when both noun *I* was high *and* the pairing was compatible. An examination of intertrial repetitions failed to find evidence of organization in recall. Where only one word from a pair was recalled it tended to come from an incompatible pair, supporting the hypothesis that the recall of one word from an incompatible pair is less likely to cue recall of the other word than is the recall of one word from a compatible pair.

Introduction

Several studies of free recall (e.g. Paivio, Yuille and Roger, 1969; Tulving, McNulty and Ozier, 1965) have shown that the image-evoking value (*I*) of nouns is related to their frequency of recall. Yuille, Paivio and Lambert (1969) demonstrated that the *I* of adjectives can be used to predict paired associate learning. The experiment described below investigated the role of the *I* of single words in the free recall of simple grammatical units. Subjects learned a list of adjective-noun pairs in which the *I* of the adjective and the nouns was varied.

The adjective and the noun forming an adjective-noun pair may or may not be compatible; that is, the resulting pair may or may not be immediately meaningful. For example, "brown" and "dog" form an immediately meaningful pair while "brown" and "afterlife" do not. The compatibility (*C*) of a pair may be expected to influence its recall since *C* may be taken to indicate the degree of association between the two words and thus the likelihood when one of the pair is recalled of it cueing the recall of the second word. Also, incompatible pairs, being meaningless should be less amenable to organization which has been shown (Mandler, 1966) to be a major determinant of the number of items free recalled. Gonzalez and Cofer (1959) found greater Ratios of Repetition, a measure of organization in free recall, for compatible pairs than for incompatible pairs. In the experiment described below the *C* of the pairs was varied to determine its affect upon recall. Compatible pairs of high *I* words are likely to evoke integrated images incorporating the features of the adjective and the noun (e.g. "brown dog" is likely to evoke an image of a brown dog). Incompatible pairs are unlikely to evoke integrated images

and pairs with words of low *I* are unlikely to evoke images at all. This suggests the possibility of an interaction between *C* and *I*; the present experiment employed a factorial design to test for such a possibility.

Method

List construction

A list of 24 adjective-noun pairs was prepared so that the C of the pairs and the I of the adjectives and the nouns was varied according to a 2 by 2 by 2 factorial design with three pairs in each of the eight subsets. The factors were adjective *I* (high or low) noun *I* (high or low) and *C* (compatible or incompatible pairing).

Nouns were selected controlling for meaningfulness (*m*) on the basis of the noun *I* ratings given by Paivio, Yuille and Madigan (1968) which have been shown to be applicable to British subjects (Morris and Reid, 1971). *I* ratings of 97 common adjectives were obtained following the procedure of Paivio *et al.* (1968). The adjectives were rated on a seven point Low Imagery—High Imagery scale by 36 undergraduates and the mean rating calculated for each word.

TABLE I

Burning village Terrified lad Yellow honeycomb	Compatible	High Adj *I*	High Noun *I*
Happy thought Old theory Haunted mind	Compatible	High Adj *I*	Low Noun *I*
Next building Big forehead Simple microscope	Compatible	Low Adj *I*	High Noun *I*
Lengthy explanation Continuous reminder Essential belief	Compatible	Low Adj *I*	Low Noun *I*
Drunken bungalow Yawning hurdle Murdered tweezers	Incompatible	High Adj *I*	High Noun *I*
White amount Wet mastery Brown deduction	Incompatible	High Adj *I*	Low Noun *I*
Slow dress Delayed chin Generous boulder	Incompatible	Low Adj *I*	High Noun *I*
Dipped simile Boxed exclusion Coded afterlife	Incompatible	Low Adj *I*	Low Noun *I*

Selecting from the highest and lowest rated adjectives and nouns 40 pairs were formed, five for each of the eight subsets of the list, which, in the judgment of the experimenters fulfilled the relevant *C* conditions. These 40 pairs were typed on cards and presented to 12 judges who were asked to sort them into two categories, one consisting of meaningful and one of meaningless pairs. On the basis of these sortings 24 pairs were selected, three for each subset of the design. Pairs unanimously sorted into one category were selected

first, then pairs placed once in the alternative category. No pair was included that had been placed in the alternative category more than twice in the 12 possible occasions. The probability of a pair being placed by chance in the same category 10 times in 12 sortings is less than 0·02. The final list is given in Table I and the ranges and means of I and m are given in Table II.

TABLE II

		I		m	
		Range	Mean	Range	Mean
Adjective	Low I	1·75–3·56	2·65	—	—
	High I	5·17–5·99	5·48	—	—
Noun	Low I	2·40–3·03	2·76	5·08–6·16	5·54
	High I	6·13–6·57	6·37	5·08–5·88	5·58

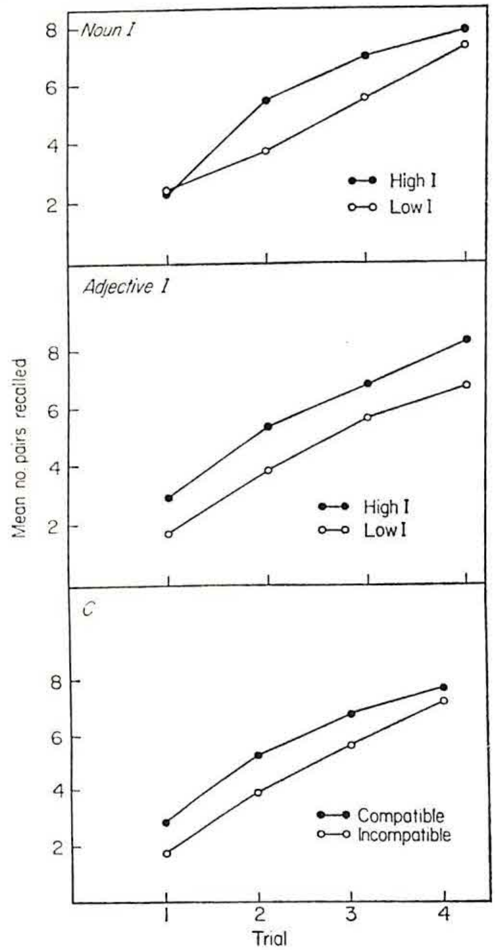

FIGURE 1. Recall per trial for the different levels of adjective I, noun I and C.

Subjects

Thirty-two undergraduates who freely agreed to take part acted as subjects.

Procedure

Each subject received four trials. Four different random orders of the list were prepared and 16 subjects received these in one order and 16 in the reverse order. The lists were presented on a Forth memory drum at a rate of one pair every 3 sec. After viewing the last pair, subjects were allowed $2\frac{1}{2}$ min to write down as many pairs as they could remember. They were told that the order in which they wrote the pairs did not matter. They were to try to maintain the pairings, but if they could not remember both words, to write down the word that they could recall. At the end of the recall period the papers were collected and the next trial began.

Results

Pair recall

The results for adjective *I*, noun *I* and *C* are shown in Figure 1. A within-subject analysis of variance (Winer, 1961) was conducted upon the data for the number of pairs correctly recalled. The four main effects of adjective *I*, noun *I*, *C* and trials were all highly significant (adjective *I*: $F\ (1, 31) = 26{\cdot}26$, $P < 0{\cdot}0001$; noun *I*: $F\ (1, 31) = 14{\cdot}5$, $P < 0{\cdot}001$; *C*: $F\ (1, 31) = 20{\cdot}94$, $P < 0{\cdot}001$; trials: $F\ (3, 93) = 202{\cdot}88$, $P < 0{\cdot}0001$). There were significant interactions between noun *I* and *C* ($F\ (1, 31) = 17{\cdot}94$, $P < 0{\cdot}001$) and between noun *I* and trials ($F\ (3, 93) = 8{\cdot}19$; $P < 0{\cdot}001$). All other interactions were nonsignificant. The interaction between noun *I* and *C* was found to result from noun *I* significantly affecting recall of compatible pairs ($F\ (1, 31) = 31{\cdot}58$, $P < 0{\cdot}0001$) but not incompatible pairs ($F\ (1, 31) = 0{\cdot}01$, $P > 0{\cdot}9$) and *C* affecting the recall of high *I* nouns ($F\ (1, 31) = 30{\cdot}02$, $P < 0{\cdot}0001$) but not low *I* nouns ($F\ (1,31) = 0{\cdot}14$, $P > 0{\cdot}7$). Figure 2 illustrates this interaction. The interaction of noun *I* with

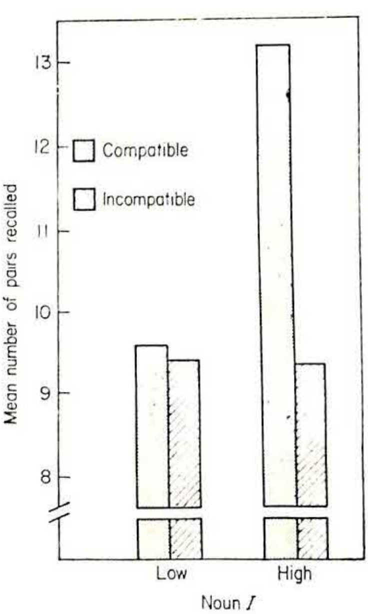

FIGURE 2. Recall of compatible and incompatible pairs with high and low noun *I*.

trials resulted from noun I having no effect on trials 1 and 4 ($P > 0 \cdot 1$) but a significant effect on trial 2 ($F(1, 31) = 27 \cdot 52$, $P < 0 \cdot 0001$) and on trial 3 ($F(1, 31) = 13 \cdot 32$, $P < 0 \cdot 001$). There was a significant improvement over trials for both high and low I nouns ($P < 0 \cdot 0001$).

Organization in recall

Each subject's recall was examined for indications of organization in recall by calculating intertrial repetitions (ITRs) between trials 3 and 4 by the method described by Bousfield and Bousfield (1966). Since it is a general finding that ITRs increase with trials it was assumed that they would appear in the final trials if at all. No evidence of intertrial organization was found since 14 subjects produced more ITRs than expected by chance (i.e. one or more) and 18 produced less (i.e. none).

Single words recalled

In a total of 194 instances only one word from a pair was recalled. With a mean of 1·5 per subject per trial this was too few for a trial by trial analysis and the data was combined across trials for each subject. Sign tests revealed that significantly more subjects recalled more high I adjectives than low I adjectives (more high I: 18; more low I: 4; $P < 0 \cdot 01$) but there was no difference for high I and low I nouns (more high I: 13; more low I: 9; $P > 0 \cdot 2$). Most subjects recalled more words from incompatible than from compatible pairs (more from compatible: 2, more from incompatible: 23; $P < 0 \cdot 001$).

Discussion

The results indicate that the I of a word, rated on its own, is related to the recall of a grammatical unit in which the word occurs. For adjectives the relationship is simple; pairs with high I adjectives are more likely to be recalled irrespective of the I of the other word or the C of the pairing. For nouns the situation is more complex. There is no obvious reason why noun I should affect recall on trials two and three but not on trials one and four. The interaction of noun I and C shows that the two variables only influence recall when both the noun I is high and the pairing is compatible. A possible explanation may be that this is the only condition in which integrated images incorporating the features of the adjective and the noun frequently occur. Bower (1970) has shown that facilitation of memory through imaging occurs only if the items to be remembered are imagined in some interactive relationship with one another. Something similar may be the case for the free recall of adjective-noun pairs where only an integrated image of adjective and noun features may serve to cue the recall of both words in the pair.

The recall of more single words from incompatible pairs fits with the hypothesis that the recall of one word from an incompatible pair is less likely to cue the recall of the other word than is the recall of one word from a compatible pair.

Tulving *et al.* (1965) found greater Subjective Organization for high I words and suggested that I may be effective in free recall because it affects the ease with which words can be grouped into higher order memory unit. Similarly, we have

argued above that the interaction of C and I may result from the forming of higher order units, in this case integrated images. Although higher order memory units probably account for some of the superiority of the high I words, this is not the whole story since Paivio *et al.* (1969) found no more ITRs for high than for low I words and the ITR data reported above was insignificant. Paivio suggests that I affects the availability of the items. High I items have two coding systems, images and a verbal form, while low I items have only a verbal coding system. If two systems in parallel are better than one, high I items should have a higher probability of recall. However, higher order memory units and greater availability do not fully account for the superiority of high I words, since in a recognition task, where availability is not in question and higher order units may even be a hinderance, subjects still perform better with high I words (Gorman, 1961). It would seem that the superiority of high I items in free recall is due not to one but to several factors.

References

BOUSFIELD, A. K. and BOUSFIELD, W. A. (1966). Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, **19**, 935–42.

BOWER, G. H. (1970). Imagery as a relational organiser in associative learning. *Journal of Verbal Learning and Verbal Behavior*, **9**, 529–33.

GONZALEZ, R. C. and COFER, C. N. (1959). Exploratory studies of verbal context by means of clustering in free recall. *Journal of Genetic Psychology*, **95**, 293–320.

GORMAN, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, **61**, 23–9.

MANDLER, G. (1966). Organization and Memory. In Spence, K. W. and Spence, J. T. (Eds.), *The Psychology of Learning and Motivation*. New York: Academic Press.

MORRIS, P. E. and REID, R. L. Canadian and British ratings of the Imagery values of words. *British Journal of Psychology*. (In press).

PAIVIO, A., YUILLE, J. C. and MADIGAN, S. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, **76**, (1, Pt. 2).

PAIVIO, A., YUILLE, J. C. and ROGERS, T. B. (1969). Noun imagery and meaningfulness in free and serial recall. *Journal of Experimental Psychology*, **79**, 509–14.

TULVING, E., MCNULTY, J. A. and OZIER, M. (1965). Vividness of words and learning to learn in free-recall learning. *Canadian Journal of Psychology*, **19**, 242–52.

WINER, B. J. (1961). *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

YUILLE, J. C., PAIVIO, A. and LAMBERT, W. E. (1969). Noun and adjective imagery and order in paired-associate learning by French and English subjects. *Canadian Journal of Psychology*, **23**, 459–66.

Received 13 July 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 399–409

SHORT TERM RETENTION OF SENTENCES†

R. A. BOAKES AND B. LODWICK‡

University of Sussex

A series of experiments was performed on the interaction between the short-term retention of sentences and of digits. In Experiment I a digit span method was used whereby subjects were presented with a sentence followed by a sequence of digits and were required either (a) to recall the sentence first and then the digits or (b) to recall the digits followed by the sentence. Under condition (a) prior recall of the sentence reduced the percentage of digit sequences correctly recalled, while under condition (b) retention of the sentence appeared to have no effect on digit recall. This last finding was confirmed in Experiment II, where the sentences varied both in grammatical complexity and length.

In Experiment III the effect of prior recall of a sentence on the recall of digits was found to depend on the type of sentence used. A correlation was observed between the size of this effect and the time taken to recall a sentence. The rate of forgetting suggested by this observation was comparable to that obtained in Experiment IV, where subjects performed an intervening task that did not involve immediate memory for sentences in the interval between the presentation and recall of a six-digit sequence.

It was concluded from these results that the short-term retention of sentences and of lists of items cannot be explained in terms of some general store of limited capacity.

Introduction

Attempts to determine the relationship between formal descriptions of language and the processes of speaking and understanding a language have been based on the assumption that a major factor determining performance is a limited capacity to store items over a short period of time. Miller and Chomsky (1965) have suggested, for example, that difficulties in understanding sentences with a number of self-embedded clauses, which are perfectly acceptable on formal grounds, reflect the limitations of such a short term store. Furthermore, it is generally assumed that this store is a general one and not specific to language. It follows that any task which involves the processing of language material should interfere with the simultaneous short term retention of other items.

An effect of this kind was found in a study by Savin and Perchonock (1965). In their experiment the subjects were presented with a sentence followed by a list of unconnected words and were asked to recall both the sentence and the words immediately afterwards. They found that the number of words recalled was related to the grammatical complexity of the sentence. The interpretation given

† Reprints may be obtained from the Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex.

‡ Now at the Chelsea College of Technology.

to these results resembles Archimedes' measurement of volume: the more complex the sentence, the more space is required to retain syntactic information and thus less is available for the extra words.

This ingenious method is of considerable interest both as a method for measuring the complexity of grammatical structures and as a way of studying the role of short memory in language performance. The study has been widely cited (e.g. Neisser, 1966) as providing evidence on the way in which grammatical complexity affects the retention of sentences. Two subsequent papers by Matthews (1968) and Wright (1968) have criticized the conclusion by Savin and Perchonock that their results were primarily determined by the transformational complexity of the sentences. An assumption made in all three papers is that, when two kinds of material are simultaneously remembered for a short period, any interference effects are to be interpreted in terms of limited storage capacity.

According to this account, when a subject learns A and B and subsequently recalls A followed by B, changes in the difficulty of recalling B reflect the difficulty of *storing* A. An alternative account is that interference occurs mainly during recall, so that recalling B is affected by the prior *recall* of A. Since it is known that even a very brief interpolated task has a drastic effect on immediate memory (e.g. Conrad, 1960), this would seem to be just as likely a basis for the interference. The following series of experiments was performed in order to determine whether interference between the retention of sentences and that of non-linguistic material is primarily a result of limited storage or of interference in recall.

Experiment I

A pilot study using the conditions described by Savin and Perchonock (1965) suggested that digits were more suitable as extra items than the list of nouns that they used. Subjects appeared to group the nouns into mnemonic phrases, whereas digits seemed more likely to be recalled as independent items. Since there have been no previous studies of interactions between the retention of sentence material and short term memory for digits, the first experiment took the form of a classical digit-span study (Woodworth and Schlosberg, 1954) which was modified by requiring the subject to recall a sentence either immediately prior to or immediately after recall of a string of digits.

Method

Material

Sequences of from 3 to 12 digits were derived from random number tables with the constraint that no digit occurred twice in succession. The sentences were of two kinds: past affirmative active, of the form "The boy has hit the ball", which for convenience will be referred to as "kernel" sentences, and negative passives, of the form "The ball has not been kicked by the boy". All sentences of the second type were obtained by transforming those of the first type. The material was recorded on tape, using a 0·8 sec/digit rate of presentation and the sentences were read with normal intonation.

Subjects

Thirty-two student volunteers from the Brighton College of Education and the University of Sussex acted as subjects, all of whom were within the age range of 18–24 years. These were randomly assigned to two groups of 16 and were individually tested.

Procedure

Each subject was given four tests. Test A consisted of a simple digit-span measurement in which the subject was required to repeat a sequence of digits immediately after the last item in the sequence had been presented. The first block of four sequences contained three digits, the next block of four sequences contained four digits and so forth. The test ended when the subject failed to recall correctly any sequence in two successive blocks. Each digit sequence was preceded by a click.

In Test B each trial consisted of the presentation of a sentence immediately followed by a sequence of digits. Subjects were instructed to repeat the sentence immediately after the last digit had been presented and then repeat the digit sequence.

For Group 1 the sentences were all of the Kernel type, while for Group 2 the sentences were all of the Negative Passive type. As in Test A the number of digits in each sequence was increased by one every block of four trials. For the first block each sequence contained four digits; in the last block each sequence contained eight digits. Thus 20 trials were given in this test.

In Test C the sentences were presented immediately prior to the digit sequences as before, but the subjects were instructed to repeat the digits before repeating the sentence. For each group the sentences were of the same type as Test B. The same procedure and number of trials were used as in Test B. Half the subjects in each group received Test B before Test C and for the other half this order was reversed.

Finally in Test D all subjects were again given a simple digit-span test in which the procedure was identical to that in Test A. This test was introduced as a means of assessing practice effects.

Results and Discussion

A subject was scored correct on any trial of Tests A and D if he correctly recalled the digit sequence with respect to its items and their order. In Tests B and C only those trials in which the sentence was recalled verbatim were considered and on these trials correct digit sequences were scored as before. The probability of correctly recalling the digit sequences in Tests A and D and the conditional probability of correct digit recall, given correct recall of the sentence, are shown in Figure 1(a) and (b). These frequencies are obtained by averaging over all subjects within a group, since both a comparison of the results for Tests A and D and examination of the results for the two halves of each group suggested that order effects were negligible. Thus each point on the solid lines in Figure 1 represents the percentage correct out of 64 trials and each point on the dashed lines represents the percentage correct out of at most 64 trials. The conditional probability for Tests B and C differed very little from the absolute probability since the sentences were incorrectly recalled on very few of the trials. The percentage of trials in which errors were made in sentence recall ranged from one per cent for the four digit blocks to nine per cent for the eight digit blocks.

There are two important aspects to these results. First there is no evidence that simple retention of a sentence interferes with performance on digit span task; the digit span measured under the conditions of Test C is no different from that under the simple conditions of Test A and D. This is immediately apparent from the results for Group 1; for Group 2 in Figure 1(b) the slight drop in the number of sequences recalled for seven-digit sequences in Test C was not found to be significant. A signs test on the differences between average scores on Tests A and D and scores on Test C for the seven-digit blocks gave $P > 0{\cdot}25$.

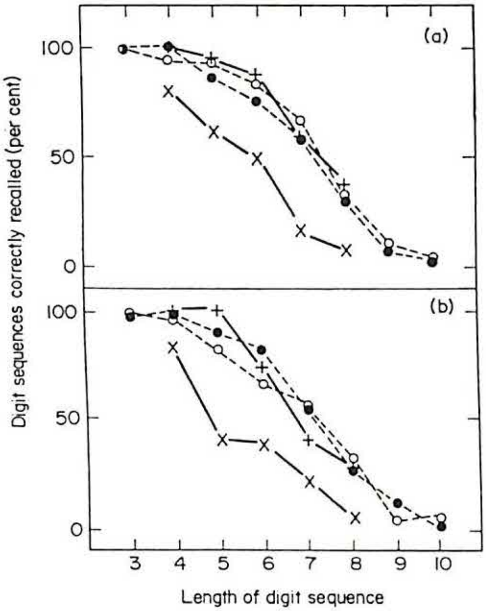

FIGURE 1. Percentage of digit sequences correctly recalled as a function of sequence length. The results are from (a) Group 1 (K sentences) and (b) Group 2 (NP sentences) in Experiment I. ● - - - - ● Test A; × —— × Test B; ○ —— ○ Test C; + - - - - + Test D.

The second aspect is the interference in Test B caused by prior recall of the sentence. In a comparison of performances on this test by Groups 1 and 2 no significant differences were detected between the amount of interference caused by the Kernel and by the negative passive sentences. One likely reason for the discrepancy between this and the results of Savin and Perchonock is that in the present experiment a subject was presented with only one form of sentence. Consequently in the following two experiments subjects were required to remember sentences of varying forms.

Experiment II

The most interesting aspect of the previous results was the absence of any interference due to the retention of the sentence. It was decided to discover whether this would also hold when the sentences presented to an individual subject varied in length and complexity and under conditions that allowed more precise measurement than the progressive digit-span test used in Experiment I. A possible criticism of this method is that any interference may be masked by a cueing effect, in that a sentence may allow more effective preparation for the digit sequence that follows it than the brief click used in Tests A and D. Experiment II was designed to remove this possible fault.

Method

Materials

A list of 50 items was prepared, each item consisting of a sentence followed by a sequence of seven digits. Six types of sentences were used. Types 1 and 2 were kernel and negative

passives as previously described (10 examples of each). Type 3 was of the form "John is eager to please" (ten examples) and Type 4 of the form "John is easy to please" (ten examples). Type 5 was of the form "The woman persuaded John to see the specialist" (five examples) and Type 6 of the form "The woman expected John to see the specialist" (five examples). Types 3 and 4 are of the same word length and contain the same number of content words as the kernel sentences. These two sets of sentences were obtained from a previous study (Blumenthal and Boakes, 1967) which had failed to detect any difference between them in terms of general difficulty. Both this pair and Types 5 and 6 have identical surface structure but differ in terms of their underlying properties (Chomsky, 1965) and thus the inclusion of these pairs permits the detection of any effects arising from these relatively subtle formal differences.

The ordering of the list was such that each successive block of ten items contained two examples each of sentence Types 1, 2, 3 and 4 and one example each of sentence Types 5 and 6. Within a block order of sentence types was randomized. Presentation rates were as in Experiment I.

Subjects

Thirty-three students of the Brighton College of Education acted as subjects and were tested in two groups, one of 18 and one of 15.

Procedure

Following the presentation of each item subjects were instructed to write down the seven digits followed by the sentence. It was stressed that they should not write down the sentence first. They were allowed a maximum of 30 sec to do this. On eight out of the 50 trials the item was preceded by instructions to ignore the sentence and merely recall the digit sequence. These instructions preceded two examples each of sentences Types 1, 2, 3 and 4. The individual sentences that were ignored in this way were different for the two groups. This procedure was introduced to provide a measure of digit recall in the absence of sentence retention, for which there was no possibility of contamination by the cueing factor discussed above.

Prior to the beginning of the experiment five practice trials were given in order to allow subjects to become familiar with the procedure.

Results

The results were scored in the same way as in Experiment I. The probability of correct recall of a sentence of a given type and the conditional probability of correct digit sequence recall, give that the sentence was correct, are shown in Table I. It can be seen that approximately half the sequences of digits are recalled regardless of whether a sentence of any type had also to be retained. This experiment thus confirms the relevant part of Experiment I for a situation in which the variety and complexity of the sentence material is very much increased. The slight differences in digit recall between Type 3 and Type 4 and also between Type 5 and Type 6 were not found to be significant when a comparison test was applied individually to these pairs ($P > 0{\cdot}10$).

Discussion

Since the completion of these experiments Epstein (1969) has reported a similar study in which the effects of reversal of order of recall were investigated. His main results are in agreement with those found here, but in addition he found a small effect of sentence retention on the recall of a list of items. The absence of any

TABLE I

Probabilities of recalling the sentence and conditional probabilities of recalling the digit sequence, given correct recall of the sentence, obtained by averaging across both groups of Experiment II (N = 33)

Sentence type	*P* (Sentence recalled)	*P* (Correct digit recall/ sentence correct)
No retention of sentence	—	0·52
Type 1 ("The boy has hit the ball")	0·91	0·50
Type 2 ("The ball has not been hit by the boy")	0·96	0·48
Type 3 ("John is eager to please")	0·93	0·50
Type 4 ("John is easy to please")	0·97	0·56
Type 5 ("The woman persuaded John to see the specialist")	0·92	0·51
Type 6 ("The woman expected John to see the specialist")	0·89	0·48

suggestion of such an effect in Experiment II is probably due to two differences between the studies: (1) Epstein used lists of words, which are thus more similar to the sentences that are digit sequences, and (2) his subjects were informed of the order of recall required on a given trial after the sentence and list had been presented.

Experiment III

It was found in Experiment I that, although requiring a subject to recall a sentence before recalling a digit sequence resulted in considerable interference to the digit task, the effect was no larger for negative passive sentences than it was for kernel sentences. In order to determine whether this discrepancy with results of Savin and Perchonock arose from the fact that a given subject was presented with sentences of only one type, as suggested in Wright (1968), a variety of sentences was presented to each subject in the present experiment. As in Experiment II all the digit sequences were of the same length with the purpose of increasing the likelihood of detecting any possible differences.

A large number of experiments, notably those following Brown (1958) and Peterson and Peterson (1959), have shown that a major factor in tasks of the present kind is the delay elapsing between presentation and recall. In view of this it is possible that the effect found by Savin and Perchonock indicates only that more complex sentences require more time for recall. This possibility was investigated in Experiment III by measuring the time taken to recall the sentence on each trial.

Method

Material

Each item consisted of a sentence followed by a sequence of six digits. A total of 50 items was used for each subject. The sentences were identical to those used in Experiment II. Six-digit sequences were used since Experiment I indicated that this length was likely to be the most suitable for detecting any differences. Two lists of items, A and B, were prepared which differed only in the pairing of digit sequences with sentences. This was done to reduce the possibility of chance associations between a given type of sentence and easy digit sequences. Both lists were tape-recorded.

Subjects

Twenty-two student volunteers from the same population as in Experiment I were used as subjects. None had been a subject in either Experiment I or II. The first 12 subjects were tested on List A and the remainder on List B. All subjects were tested individually.

Procedure

The procedure was similar to that for Test B in Experiment I. Following the presentation of an item subjects were required to recall the sentence and then the digit sequence. Their responses were recorded on tape, since it seemed likely that the inter-subject variability in the time taken to write down a sentence would be much greater than that of the time to say it aloud. Subsequently the time for each trial between the end of the presentation of an item and the end of the recall of sentence was estimated by an assistant using a stop-watch. The assistant was not informed of the purpose of the experiment. Following three or four practice trials the 50 test items were presented in a single session with an inter-trial interval of approximately 30 sec.

Results

For each type of sentence the probability of correct recall of a sentence, the conditional probability of correct recall of the digit sequences, given the sentence correct, and the average times for recalling a sentence are shown in Table II. Since there were no systematic differences between the two groups these figures represent means obtained from all 22 subjects. The average times for recalling a sentence were based only on trials in which the sentences were correctly recalled. The effects of sentence type both on the probability of correct digit recall and on the time of recall were found to be significant (F (5, 105), $P < 0{\cdot}01$). Over sentence types the rank correlation between digit recall and sentence recall time was $0{\cdot}94$ ($P < 0{\cdot}01$).

A comparison with other studies can be made by examining digit recall following sentences of either Type 1 (kernel) or Type 2 (negative passive). The mean number of words recalled following a kernel sentence was 5·3 in Savin and Perchonock (1965) and 3·6 in Wright (1968); in the present experiment a mean of 4·9 digits were recalled after kernel sentences. Similarly the comparable figures for negative passive sentences are 3·5 (Savin and Perchonock), 3·2 (Wright) and 3·9 digits in this experiment. Thus the use of digit sequences does not appear to have introduced any greater changes in the size of the effects obtained than between the two earlier studies that used very similar material.

The claim by Matthews (1968) that the transformational complexity of a sentence is not a determining factor in this kind of situation, whereas its length is, was examined by comparing performance with Type 1 sentences and that with

TABLE II

Probability of sentence recall, conditional probability of digit sequence recall and average times for recalling sentence in Experiment III

Sentence type	P (sentence correct)	P (digits correct/ sentence correct)	Average time for sentence recall
Type 1 ("The boy has hit the ball")	0·96	0·39	3·0 sec
Type 2 ("The ball has not been hit by the boy")	0·97	0·27	3·8 sec
Type 3 ("John is eager to please")	0·98	0·29	3·1 sec
Type 4 ("John is easy to please")	0·92	0·30	3·3 sec
Type 5 ("The woman persuaded John to see the specialist")	0·85	0·27	3·9 sec
Type 6 ("The woman expected John to see the specialist")	0·86	0·19	4·1 sec

sentences of Types 3 and 4. All three types were equivalent in terms of number of content words and average number of syllables, while the number of words was one fewer for Types 3 and 4. The probability of recalling a digit sequence was greater ($P < 0{\cdot}05$) following a Type 1 sentence than following either Types 3 or 4. This difference would seem to be a result of the greater structural complexity of sentences like "John is eager (or easy) to please". However the relatively subtle syntactic differences between Types 3 and 4 and between Types 5 and 6 did not produce any significant differences in the probability of recalling the digit sequence that followed.

Discussion

These results show that the effect discovered by Savin and Perchonock can also be obtained with the material used in the present set of experiments. The hypothesis that this effect is due to interference during recall is supported by the observed relationship between the time taken to recall a given type of sentence and subsequent performance in recalling the string of digits.

The rate of forgetting suggested by these results is faster than that typically found in studies following Peterson and Peterson (1959): a six-digit sequence was correctly recalled on about 80% of the trials in Experiment I, while the results from Experiment III suggest that a delay of 3 sec is sufficient to reduce this figure to 40%. This fast forgetting rate might well be due to the unusual length of the sequences used in the present task and, since no exactly comparable data exists for the short-term retention of digit sequences of this length, the following experiment was designed to provide this information.

Experiment IV

Method

Material

Sequences of six digits were used as in Experiment III.

Subjects

Twelve subjects from the same population as Experiment I. None had taken part in any of the previous experiments.

Procedure

On each trial a digit sequence was read out at the rate of 0·8 sec/digit, followed by "Now" or a letter of the alphabet chosen at random. Subjects were instructed to recall the sequence immediately if the experimenter said "Now". If a letter was given, the instructions were to recite the alphabet as rapidly as possible, beginning at that letter. When the experimenter subsequently said "Now", the digit sequence was to be recalled.

Following six practice trials a total of 30 trials was given. On ten trials "Now" occurred immediately after the presentation of the digits, on ten trials 2·5 sec after the letter and on ten trials 5 sec after the letter. The sequence of delay intervals was randomized and changed for each subject. Subjects were tested individually and their verbal recall was recorded by the experimenter on a score sheet.

Results

The length of the intervening interval had a highly reliable effect on the percentage of digit sequences correctly recalled ($P < 0{\cdot}001$, Friedman 2-way analysis of variance). In Figure 2 the percentage of sequences correctly recalled is plotted as a function of the intervening interval. Also shown in this figure are the results from Experiment III, where for each sentence type the percentage of digit sequences correctly recalled is plotted as a function of the average time taken to recall this type of sentence.

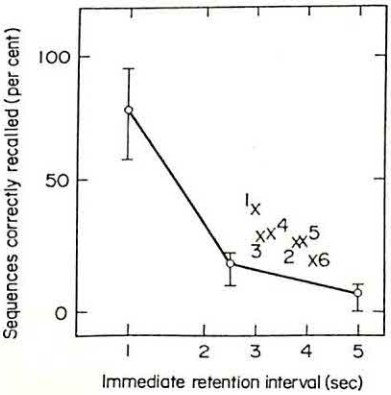

FIGURE 2. Percentage of digit sequences correctly recalled as a function of the retention interval in Experiment IV. The circles represent these percentages, averaged over subjects, and also indicated are the inter-quartile ranges. Results from Experiment III are shown by crosses; each cross indicates the percentage recall of a subsequent digit sequence and the mean time for sentence recall for a given sentence type.

X from Experiment III; $\overset{\top}{\underset{\bot}{\text{o}}}$ from Experiment IV.

Exact quantitative comparisons cannot be made in view of the necessarily different ways of estimating temporal intervals in the two experiments. Nevertheless it is clear that, for example, a 3 sec recital of the alphabet produces the same order of forgetting as 3 sec devoted to recalling a sentence.

Discussion

Both Experiments I and II indicate that the short term *retention* of a sentence has no effect on the simultaneous retention and recall of a sequence of digits. In contrast Experiment III shows that, with the same set of sentences, interpolating *recall* of a sentence between the presentation and recall of a digit sequence does interfere with the recall of the digits. It is therefore clear that the kind of interference obtained when subjects are required to remember both a sentence and some other material cannot be interpreted in terms of a limited amount of space being available in immediate memory.

The results of Experiment IV show that even a very brief interpolation of some other task, which does not involve immediate memory, leads to a large decrement in the recall of a long digit sequence. It therefore seems plausible that the kind of interference obtained in Experiment III arose, not because the sentences had to be remembered for a short time, but because recall of a sentence imposes a delay, in which no rehearsal can occur, on the recall of the digits. Performances on the digit task is thus mainly an indirect measure of how long it takes to recall a sentence. This appears to be determined both by the length and the complexity of the sentence. This conclusion is supported by the results of Glucksberg and Danks (1969), who found a similar relationship between time and recall performance in a study closely following that of Savon and Perchonock (1965).

In a general sense, it is clear that a major factor in the discrepancy between linguistic competence and actual performance is that of memory limitations over short periods of time. A long and complex sentence, that is grammatically acceptable and quite easy to understand in written form, may well be incomprehensible when spoken. However, as indicated earlier, many discussions of this point make further, often implicit, assumptions. These are, first, that the phenomena studied in short-term verbal memory research directly reflect processes whose primary function is a linguistic one, and, second, that these phenomena are most appropriately explained in terms of limited capacity mechanisms.

This second assumption has been a very resilient one. Yet many of the phenomena that directly suggest this idea, including that of the constant memory span, are explicable in terms of models that contain no assumption of limited storage (e.g. Laughery, 1969). The present results indicate that it is not an appropriate way of accounting for the way that retaining and recalling a sentence interferes with a digit task. Furthermore, the finding that retaining a fairly long and complex sentence for a short time has no effect on the simultaneous retention and recall of a digit sequence throws doubt on the first assumption. There may well be no close relationship between short-term memory mechanisms involved in the processing of sentences and the performance of subjects memorizing lists of unconnected items.

We are very grateful for the assistance of Miss H. Rosenberg and T. J. Lobstein.

References

BLUMENTHAL, A. L. and BOAKES, R. A. (1967). Prompted recall of sentences. *Journal of Verbal Learning and Verbal Behavior*, **6**, 674–6.

BROWN, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, **10**, 12–21.

CHOMSKY, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT.

CONRAD, R. (1960). Very brief delay of immediate recall. *Quarterly Journal of Experimental Psychology*, **12**, 45–7.

EPSTEIN, W. (1969). Recall of word lists following learning of sentences and of anomalous and random strings. *Journal of Verbal Learning and Verbal Behavior*, **8**, 20–5.

GLUCKSBERG, S. and DANKS, J. H. (1969). Grammatical structure and recall: a function of the space in immediate memory or a recall delay? *Perception and Psychophysics*, **6**, 113–7.

LAUGHERY, K. R. (1969). Computer simulation of short-term memory: a component decay model. *Psychology of Learning and Motivation*, **3**, 135–97.

MATTHEWS, W. A. (1968). Transformational complexity and short term recall. *Language and Speech*, **11**, 120–8.

MILLER, G. A. and CHOMSKY, N. (1965). Finitary models of language users. In LUCE, R. D., BUSH, R. R. and GALANTER, E. (Eds.), *Handbook of Mathematical Psychology, Vol*. 2. New York: Wiley.

NEISSER, H. (1966). *Cognitive Psychology*. Pp. 272–6. New York: Appleton-Century-Crofts.

PETERSON, L. R. and PETERSON, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, **58**, 193–8.

SAVIN, H. B. and PERCHONOCK, E. (1965). Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, **4**, 348–53.

WOODWORTH, R. S. and SCHLOSBERG, H. (1954). *Experimental Psychology: Revised Edition*. New York: Holt-Rinehart-Winston.

WRIGHT, P. (1968). Sentence retention and transformation theory. *Quarterly Journal of Experimental Psychology*, **20**, 265–72.

Received 11 *March* 1971

Quarterly Journal of Experimental Psychology (1971) 23, 410–416

LATERAL WORD RECOGNITION: EFFECTS OF UNILATERAL AND BILATERAL PRESENTATION, ASYNCHRONY OF BILATERAL PRESENTATION, AND FORCED ORDER OF REPORT

WALTER F. MCKEEVER

Bowling Green State University, Bowling Green, Ohio, U.S.A.

This experiment enquired: (1) whether right visual field (RVF) recognition superiority was greater in bilateral than in unilateral word presentation; (2) whether left field-favouring attentional or recall sets could be induced by presenting left visual field (LVF) words 20 msec prior to RVF words or by instructions to report LVF words prior to RVF words. Results showed: (1) all conditions studied yielded significant RVF superiority; (2) RVF superiority magnitude was significantly greater in bilateral than in unilateral presentation, suggesting the tenability of hypotheses that different mechanisms operate in these conditions; (3) neither earlier delivery nor earlier report of LVF words altered the pattern of RVF superiority in bilateral presentation, the later result demonstrating that differential receptive organization rather than differential recall of the two stimuli is responsible for RVF superiority in bilateral presentation.

Introduction

Sperry (1968) has shown that human commissurotomy patients are unable to report verbally words tachistoscopically presented in the left visual field (LVF), but readily report words shown in the right visual field (RVF). This inability to report LVF stimuli appears to reflect the lack of language output capabilities of the right cerebral hemisphere. Barton, Goodglass, and Shai (1965), Kimura (1966), Zurif and Bryden (1969) and others have provided data suggesting that such RVF recognition superiority in normal subjects may similarly reflect the language specialization of the left hemisphere.

A more traditional view ascribed RVF superiority to "attentional scanning" mechanisms acquired through reading experience. Central to this view was the finding (Crosland, 1931; Heron, 1957) that RVF superiority obtained only when letters were unilaterally presented in relation to the fixation point and that LVF superiority obtained with bilateral presentation. The explanation offered (Heron, 1957) was that readers of English acquire tendencies to attend initially to the left-most element of horizontal letter arrays and subsequently, from that element, to the right. Thus, White (1969), in a review of the area concluded: "... when verbal stimuli are presented bilaterally about the centre of the visual field, a superiority favouring elements in the LVF is usually found. For the sake of complete ade-

quacy, a dominance hypothesis should account for these data—which it seemingly cannot do." (p. 402.)

McKeever and Huling (1971*a*) pointed out, however, that the experiments supportive of this view had failed to control the subject's fixation during the bilateral stimulus presentation. They suggested that subjects in this situation might focus to the left of fixation in accordance with probable left-to-right reading and report sets to maximize their performance. White too (1969), noted that fixation control is required to assure that stimuli are channeled only to the intended hemispheres and that comparable degrees of left and right stimulus lateralization are maintained. The technique employed by McKeever and Huling (1971*a*) required subjects to report single digits which appeared at the fixation point on each bilateral stimulus presentation.

When fixation was thus controlled McKeever and Huling (1971*a*) found, in accordance with their hypotheses, that bilateral presentation yielded both a significant RVF superiority and an apparently greater magnitude of such superiority than they had obtained in unilateral presentation experiments (McKeever and Huling, 1970*a*). The finding of an apparently more clear-cut RVF superiority in bilateral than in unilateral presentation was consistent with their view that some central competitional mechanism would operate in bilateral presentation to enhance RVF superiority beyond the level obtained in unilateral presentation. This mechanism was postulated to be a disparity of arrival time at left hemisphere language centres, with the earlier arriving RVF stimuli creating some sort of psychological refractoriness to the subsequently arriving LVF stimuli. A direct test of this temporal disparity hypothesis (McKeever and Huling, 1971*b*) however, provided no support whatever for it.

In view of the failure to substantiate the temporal disparity hypothesis, the prior question of the extent of left-right recognition differences in the two types of presentation calls for a rigorous evaluation. It must be noted that the conclusion that the two conditions yielded different levels of RVF superiority was based on comparisons between experiments which actually differed with respect to exposure times, words, maximal scores, and in some instances, viewing conditions (monocular versus binocular). Additionally, it could be argued that when words are bilaterally presented at very brief exposures the extent of RVF superiority might be augmented over the level obtained in unilateral presentation because the task difficulty is increased to a point where subjects quickly realize the improbability of recognizing both words. They might then choose to ignore the more difficult LVF words and "concentrate" on the earlier RVF words. Such a basis of a greater RVF superiority in bilateral than unilateral presentation would be quite different than the neuropsychological bases proposed by McKeever and Huling (1971*a*, *b*).

The present experiment thus addressed the following questions: (1) Do unilateral and bilateral presentation yield different levels of RVF superiority when other major methodological factors are as comparable as possible? (2) Can attentional focusing on LVF words, induced by instructions always to report them prior to RVF words, decrease the level of RVF superiority in the bilateral presentation experiment? (3) Can the level of RVF superiority be reduced by presenting LVF

words 20 msec earlier than RVF words in bilateral presentation? This last question was raised by some data (McKeever and Huling, 1971*b*) suggesting that consistent leading of the LVF words relative to RVF words might induce subjects to "attend" to the LVF words with the result that RVF recognitions would be reduced. The question merits examination because "attentional" factors should be unimportant if performance is simply a function of which cerebral hemisphere receives the word stimulation.

Method

Apparatus

Characteristics of the Dodge-type tachistoscope employed have been fully described previously (McKeever and Huling, 1971*b*).

Stimuli

Two sets of stimulus cards were prepared, one set of 20 bilateral words and one set of 40 unilateral words. Word pairs for the bilateral set were: *bear-mask*, *mask-bear*, *cake-lane*, *lane-cake*, *cold-road*, *road-cold*, *comb-hand*, *hand-comb*, *dove-hare*, *hare-dove*, *farm-post*, *post-farm*, *knob-tape*, *tape-knob*, *lamp-snow*, *snow-lamp*, *list-shoe*, *shoe-list*, *nose-rain*, *rain-nose*. Unilateral stimuli consisted of the same 20 words, but with only one word to a card.

Stimulus cards were made with press-on letters applied to a semi-transparent, light-diffusing celluloid material. The near-point of each word to fixation corresponded to 1·6° of horizontal visual angle and each word subtended 2·4°. A single arabic number, from 2 to 9, was placed at the fixation point of each card. These were of the same size as the individual letters used in the words. The pre-exposure field was a white card with a small circle, designating the fixation point at mid-field; the circle was just large enough to be filled by the fixation number on each card as it was flashed.

Subjects and conditions

Subjects were 40 right-handed, college students free of uncorrected visual anomalies. Each of four groups of 10 subjects received a different experimental eondition. One group received unilateral presentations (UNI); a second group received simultaneous bilateral presentations (SBI); a third received bilateral presentations in which LVF words were presented 20 msec earlier than RVF words on all trials (BILFL); and the fourth group received simultaneous bilateral presentations but with the instruction to report LVF stimuli prior to reporting RVF stimuli (BILOR).

Procedure

Subjects were told that 2 sec before each stimulus presentation the experimenter would say "focus" as a signal to focus binocularly in the "focus circle". The importance of always reporting the focus number correctly and prior to anything else was explained. Two trials with a card having only a fixation number were given to accustom the subject to the procedure. Instructions from this point depended on the condition to which the subject was assigned. Subjects in the UNI condition were told: "In addition to the focus circle number each card will have a word to one side or to the other side of the number. After you report the number tell me anything you can about the word on the card. Because the exposures are brief, you may never feel 100% certain as to what the word was, so you have to feel free to venture hunches or to say 'It could have been a certain word'. Do you understand? Let's begin."

Instructions for the SBI subjects were: "In addition to the focus circle number each card will have a word or words on it. After you report the number tell me anything you can about the words or word on the card." The remainder of the instruction was identical to that for UNI subjects except for the substitution of "word or words" in place of "word". Subjects in the BILFL group received identical instructions to those in the SBI group.

Subjects in the BILOR condition were instructed: "In addition to the focus circle number each card will have two words on it, one word to the left and one to the right of the number. The two words will be different words. After you report the number tell me anything you can about the word to the left of the number and then anything you can about the word to the right of the number." The remainder of the instruction was identical to that of the SBI condition except that the experimenter reiterated at the end of the instruction the requirement to report first the number, then the LVF word, and then the RVF word.

Exposure time in all conditions was 20 msec for each field and both fields were illuminated on all trials. Subjects were never informed as to the correctness of their responses except that errors in reporting the fixation number were called to their attention in a mildly disapproving manner and they were urged to "always get the number right". The inter-stimulus interval was approximately 10 sec. Stimulus presentation order was identical in all bilateral conditions and all subjects within any condition viewed the stimuli in the same order. Scores were the number of wholly correct recognition reports. Words identified without correct report of the fixation numbers were scored as errors since subjects could have looked directly at the words rather than in the fixation circle. For all groups, the maximum score was 20 for each field.

Results

Recognitions in the RVF substantially outnumbered recognitions in the LVF in all conditions and recognitions were consistently highest in the UNI condition. Every subject showed more RVF recognitions than LVF recognitions, except for one subject in the SBI group and one subject in the BILOR group. Both of these subjects showed equal RVF and LVF scores.

Mean recognition scores for the LVF and RVF, as well as mean RVF–LVF recognition score differences within each condition are shown in Table I. These left-right differences were significant in each condition as assessed by t-ratio tests for differences between correlated means. Thus, each condition was characterized by significant RVF superiority.

TABLE I

Mean recognition scores for LVF and RVF, differences between LVF and RVF, and significance of differences for all conditions

Condition	LVF Score	RVF Score	Difference score	t	P
UNI	9·4	15·8	6·4	7·02	<0·001
SBI	2·9	11·1	8·2	4·72	<0·005
BILOR	2·5	7·6	5·1	4·87	<0·001
BILFL	3·6	12·7	9·1	8·27	<0·001

Data comparing the effects of the various conditions on LVF, RVF and total recognitions are presented in Table II. It can be seen that the UNI group made significantly more recognitions in the LVF and more total recognitions than any of the bilateral groups and significantly more RVF recognitions than the SBI and BILOR groups. The UNI condition also produced more RVF recognitions than the BILFL condition but the difference was short of statistical significance. The only significant bilateral group difference was the greater RVF recognition of the BILFL, as opposed to the BILOR, group.

TABLE II

Intergroup difference scores for RVF, LVF and total recognitions and significance of differences

Groups compared	LVF score difference	RVF score difference	Total score difference
UNI-SBI	6·5†	4·7*	11·2†
UNI-BILOR	6·9†	8·2†	15·1†
UNI-BILFL	5·8*	3·1	8·9*
SBI-BILOR	0·4	3·5	3·9
SBI-BILFL	0·7	1·6	2·3
BILOR-BILFL	1·1	5·1*	6·2

* < 0·5; † < 0·01.

To evaluate whether bilateral presentation yielded a more clear-cut RVF superiority than unilateral presentation, differences in the proportion of RVF recognitions within the experimental groups were made by means of chi square tests for differences between uncorrelated proportions. The results showed that, in comparison with the UNI condition, a significantly greater proportion of RVF recognition occurred in the SBI condition ($\chi^2 = 11·49$, $P < 0·001$), the BILOR condition ($\chi^2 = 5·06$, $P < 0·05$), and the BILFL condition ($\chi^2 = 8·01$, $P < 0·01$). Differences between bilateral conditions did not approach significance. These results show that RVF superiority was, indeed, more clear-cut in bilateral than in unilateral presentation.

The fact that unilateral presentation subjects received more trails than bilateral presentation subjects might be thought responsible for the higher level of recognition in the unilateral condition, i.e. one could argue that a practice effect would favour recognitions in this group. However, a comparison between performance on the first and second 20 trials of this group showed that on the first 20 trials there were 78 RVF and 47 LVF recognitions; on the second 20 trials there were 80 RVF and 44 LVF recognitions. It is evident from these virtually identical scores that "practice" did not account for the greater recognition nor for the pattern of recognition of the UNI subjects.

Discussion

The results of this experiment clearly demonstrate that unilateral presentation yields a greater number of recognitions than does any of the bilateral presentation conditions studied. Although unilateral presentation yielded a significant RVF superiority the proportion of RVF recognitions within this condition was significantly smaller than in any of the bilateral conditions. Bilateral presentation does, then, as previous unilateral and bilateral experiments of McKeever and Huling had suggested, yield a more clear-cut RVF superiority than does bilateral presentation. The ratios of RVF to LVF recognitions in the present experiment (1·6 to 1 in UNI; 4 to 1 in SBI) are quite close to those obtained in the separate unilateral (approximately 1·5 to 1) and bilateral (approximately 5 to 1) experiments.

The results also showed that neither leading of LVF words nor instructions to report the LVF stimuli prior to RVF stimuli had any effect on the outcome. The result for the BILFL condition resolves the unanswered question raised by the finding (McKeever and Huling, 1971*b*) of significantly lower RVF recognition in a group of subjects administered a predominance of LVF-led trials. Present findings suggest that result would best be considered a chance finding, which was one alternative originally considered by the investigators. In the present experiment, consistent leading of the LVF actually produced higher recognitions in both fields than SBI, a tendency which might suggest that this non-simultaneous presentation condition tends, however slightly, in the direction of approximating UNI conditions.

That an order of report bias favouring the RVF cannot account for either the fact or the magnitude of RVF superiority is unequivocally clear from present findings. The BILOR group showed both significant RVF superiority and a magnitude of RVF superiority that did not differ from that of the SBI group. This was true despite the consistently applied requirement that subjects report anything they could about the LVF words prior to anything they could about RVF words. The typical pattern of performance was for subjects to say something like "I don't know about the one on the left . . ." or "There was a (letter) on the left and the word (correct) on the right".

The inability of order of report instructions to influence recognition performance contrasts with the effect of order of report instructions on dichotic listening performance. Cooper, Achenbach, Satz, and Levy (1967) showed that, in addition to a right ear superiority, a significant order of report effect exists in dichotic listening performances. This suggests that while dichotic listening task performances are at least partly a function of differential *recall* of left and right ear stimuli RVF superiority in the bilateral presentation experiment does not appear to involve differential short term memory processes for left and right field words. The visual task results appear to reflect differential "registration" or receptive organization rather than memory factors. Since both dichotic listening and lateral word recognition have been posited to reflect the lateral language dominance of the left hemisphere, this difference may be noteworthy.

Finally, although the differences in recognition associated with unilateral and simultaneous bilateral presentations are consistent with the McKeever and Huling (1971*b*) hypothesis of a central competitional determinant of the greater RVF superiority in bilateral than in unilateral presentation, more definitive support for that hypothesis will require a demonstration that the difference is, at least partly, a function of the simultaneous bilaterality of presentation and not simply of the greater quantity of stimulus material in bilateral presentation.

References

Cooper, A., Achenbach, K., Satz, P. and Levy, M. (1967). Order of report and ear asymmetry in dichotic listening. *Psychonomic Science*, **29**, 97–8.

Crosland, H. R. (1931). Letter-position effects in the range of attention experiment as affected by the number of letters in each exposure. *Journal of Experimental Psychology*, **14**, 477–507.

HERON, W. (1957). Perception as a function of retinal locus and attention. *American Journal of Psychology*, **70**, 38–48.
KIMURA, D. (1966). Dual functional asymmetry of the brain in visual perception. *Neuropsychologia*, **4**, 275–85.
MCKEEVER, W. F. and HULING, M. D. (1970*a*). Left cerebral hemisphere superiority in tachistoscopic word recognition performances. *Perceptual and Motor Skills*, **30**, 763–6.
MCKEEVER, W. F. and HULING, M. D. (1970*b*). Lateral dominance in tachistoscopic word recognitions of children at two levels of ability. *Quarterly Journal of Experimental Psychology*, **22**, 600–4.
MCKEEVER, W. F. and HULING, M. D. (1971*a*). Lateral dominance in tachistoscopic word recognition performances obtained with simultaneous bilateral input. *Neuropsychologia*, **9**, 15–20.
MCKEEVER, W. F. and HULING, M. D. (1971*b*). Bilateral tachistoscopic word recognition as a function of hemisphere stimulated and interhemispheric transfer time. *Neuropsychologia* (In press).
SPERRY, R. W. (1968). Mental unity following surgical disconnection of the cerebral hemispheres. *The Harvey Lectures, Series* 62. New York: Academic Press.
WHITE, M. (1969). Laterality differences in perception: A review. *Psychological Bulletin*, **72**, 387–405.
ZURIF, E. B. and BRYDEN, M. P. (1969). Familial handedness and left-right differences in auditory and visual perception. *Neuropsychologia*, **7**, 179–87.

Received 18 *June* 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 417–422

DISCOVERY AS A MEANS TO UNDERSTANDING

PAOLO LEGRENZI†

Department of Psychology,
Istituto Superiore di Scienze Sociali, Trento, Italy

Two experiments were conducted to determine whether the discovery of a conditional rule from its instances facilitated a test of its truth from partially concealed information, compared with a situation in which the rule was presented directly. In the first experiment the subjects who discovered the rule expressed it in their own words. The results showed that they tested it significantly better than a control group who were presented with the rule, as expressed by the subjects in the experimental group. In the second experiment the subjects in the experimental group were compelled to express the rule in the form of a conditional sentence. The results showed again that they performed significantly better than those in a corresponding control group.

Introduction

Previous experiments (Wason, 1968, 1969*a*) have established that it is very difficult to decide what information is required to test the truth of an abstract conditional sentence. For example, given the sentence: *Every card which has a C on one side has a 4 on the other side* (and knowledge that each card has a letter on one side and a number on the other side), together with four cards showing respectively *C*, *F*, 4, 9, hardly any individuals make the correct choice of cards to turn over (*C* and 9) in order to determine if the sentence is true or false. This problem is called the "selection task" and the conditional sentence is called the "rule". Previous experiments have shown that intelligent adults usually choose the cards mentioned in the antecedent and the consequent of the rule (*C* and 4), or only that mentioned in the antecedent (*C*). They nearly always fail to choose the card which negates the consequent (9). However, it has been shown (Wason and Shapiro, 1971; Johnson-Laird, Legrenzi and Legrenzi, unpublished) that if this task is presented in a familiar guise, the errors are dramatically reduced.

The present investigation seeks to determine, not whether variation in the content of the task (abstract vs. concrete) affects the ease of solution, but whether the discovery of the rule by the subject facilitates its understanding, and hence diminishes the typical errors of the selection task.

Experiment I

Aim

This experiment compares the difference between the ability to test a rule, (1) after it has been discovered and formulated, and (2) when it is presented directly.

† This research was carried out while the author was resident at the Psychology Department, University College, London.

It was predicted that when the rule has been discovered and formulated, subjects will perform better in choosing the conditions which would test its truth.

Method

Design

Two independent groups were tested. The "discovery group" was just presented with the four instances (cards) of a conditional rule (three positive and one negative) illustrated in upper Figure 1, and then had to deduce the rule, and express it in words.

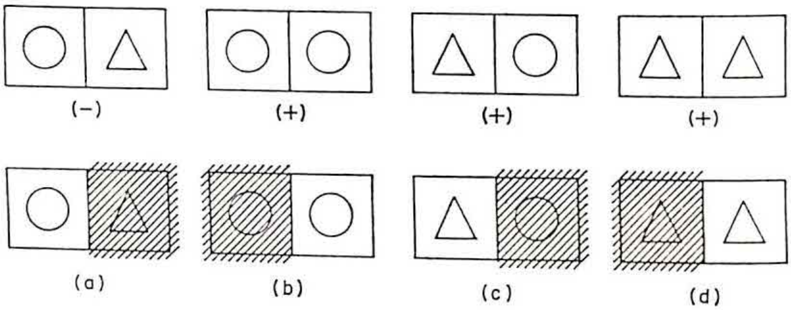

FIGURE 1. The four positive (+) and negative (−) instances of the rule, and the partially masked instances.

Four more partially masked instances (lower Figure 1) were then presented, and the task was to state which cards would need to be unmasked to find out whether the rule was true or false. The rule can be expressed in several different ways, but the conditional formulation is as follows: *If there is a circle on the left, then there is a circle on the right*. The correct solution is to choose cards (a) and (d). The control group was immediately presented with the rule as expressed by the discovery group and with the four partially masked instances, and was instructed to perform the selection task. Each expression of the rule was associated with the same number of subjects in each group. For example, if expression x was derived by, say, four subjects in the discovery group, then four subjects would be presented with the rule in that expression in the control group.

Subjects

Sixty undergraduates (paid volunteers) of University College, London, who had no previous experience of this type of task, were allocated alternately to the groups and tested individually.

Procedure

The subjects in the discovery group were instructed as follows: "You can see that I have some cards in my hands. These cards are labelled with two figures: one on the left and one on the right. The figures are triangles and circles. The way in which the triangles and the circles are arranged follows a rule. Some of the cards follow the rule but others don't. Now I will show you these cards and for each of them I shall tell you whether the rule is true or false. When you think you have found the rule, please write it down on this paper. Have you any questions? I want you to be quite clear from the start about what you have to do." The subject was then presented with the four cards (Fig. 1) and the experimenter pointed out the three which conformed to the rule, and the one which did not conform to it. When the subject had discovered the rule and expressed it in words, he was instructed as follows: "Now that you have found the rule, I shall show you some cards with the figures arranged in the same way as before, but this time half of each card will be covered. What I want you to tell me this time is whether you need to know what is on the hidden part of each card in order to decide whether the rule you have discovered is true or false. You will

have all the time you need to think about it. Have you any questions? I want you to be quite clear from the start about what you have to do."

In the control group the subjects were told: "I have four cards here. Each card has two shapes on it, one on the left and one on the right. The shapes can be triangles or circles. No other shapes are involved. You will see that one half of each of the cards is hidden from view. You know that a triangle or a circle is concealed but you don't know which. Now I am going to present you with a rule which refers to just the four cards in front of you. Your task is to decide upon those cards, and only those cards, that you would need to uncover in order to be certain whether the rule is true or false."

Results

From a logical point of view three basic types of rule were obtained: disjunctive (mentioning only all the positive instances), negative (excluding the one negative instance), and, of course, the conditional expression. The frequency with which the following rules were derived is listed below. The first number, in parentheses, after each rule, indicates the total number of subjects in the discovery group (who derived it) and in the control group (who were presented with it). The second and third numbers show respectively how many subjects in the discovery and control groups solved the selection task correctly.

Two same figures, if different a triangle on the left and a circle on the right (12) (10) (5)

Two circles, two triangles, or a triangle on the left and a circle on the right (6) (6) (4)

The rule is true if, there is a circle on the right or two triangles together (4) (4) (2)

There is not a circle on the left and a triangle on the right (3) (3) (2)

If there is a circle on the left, there is not a triangle on the right (2) (1) (1)

If a circle is together with a triangle, the circle is on the right and the triangle on the left (2) (1) (0)

If there is a circle on the left then there is a circle on the right (1) (1) (0)

It will be noted that only one subject expressed the rule as a conditional and that the most frequent expression was a disjunctive. This is consistent with the results obtained by Johnson-Laird and Hayes (see Wason and Johnson-Laird, 1972). It will also be noted that, logically, each of the expressions of the rule is correct. Two subjects were unable to express the rule in any form and they were replaced by two extra subjects in order to bring the total number of subjects to thirty.

TABLE I

Selection

	Right	Wrong	Total
Discovery group	26	4	30
Control group	14	16	30
Total	40	20	—

The prediction that the discovery group would test the rule more efficiently than the control group was confirmed as shown in Table I ($P < 0{\cdot}01$, one-tailed, Fisher–Yates exact test).

This experiment, however, is clearly not an adequate test of a conditional rule because only one subject formulated the rule as such. A way had to be found to allow the discovery of the rule, but to force it in the logical form of the conditional.

Experiment II

Aim

It was predicted that the discovery of a conditional rule, expressed as a conditional, enables subjects to test it more efficiently than when it is presented directly.

Method

Design

A discovery group was presented with the outline of a conditional expression which had to be completed when the rule had been discovered. The control group was presented with the rule directly in a conditional form. Both groups then had to perform the selection task, as in Experiment I, and the material used was the same.

Subjects

Twenty-four undergraduates (paid volunteers) of University College, London, who had no previous experience of this type of task, were allocated alternately to the groups, and tested individually.

Procedure

The instructions for the discovery group were the same as in Experiment I, but instead of asking the subjects to formulate the rule in their own words, they were asked to complete the following expression by writing down the appropriate words on the dotted lines: *If there is . . . on the left, then there is . . . on the right.* They were then instructed to perform the selection task with the partially masked cards (see Fig. 1, lower half). In the control group the subjects were instructed similarly to perform the selection task with reference to the following rule: *If there is a circle on the left, then there is a circle on the right.*

Results

Table II shows the number of subjects in each group who made the correct selection.

TABLE II

Selection

	Right	Wrong	Total
Discovery group	10	1	11
Control group	2	9	11
Total	12	10	—

The prediction was confirmed ($P < 0{\cdot}0025$, one-tailed, Fisher–Yates exact test). However, one subject in the discovery group completed the rule as follows:

If there is a triangle on the left then there is a triangle or a circle on the right. His subsequent selection was incorrect, and the corresponding subject in the control group also made an incorrect selection. All the other subjects in the discovery group completed the rule in the correct way. The results show much more clearly than the results of Experiment I that the discovery of a conditional rule, expressed in conditional form, facilitates the correct choice of instances to test it.

Discussion

Previous research has shown that the very great difficulty of the selection task with abstract material is relatively reduced when the material used is concrete, perhaps because it enables visual imagery to assist the reasoning process (Wason and Shapiro, 1971), and that it is even more strongly reduced when the task simulates exactly a real life activity (Johnson-Laird, Legrenzi and Legrenzi, unpublished).

Wason and Shapiro, however, also tried to utilize what they call "contrived experience". This was done by making the subjects familiar with the other side of the cards before the selection task is performed. The effects of this experience on the subsequent selection task were not very impressive, even when the subjects had been engaged for 24 trials in constructing positive and negative instances of the rule, by naming symbols on the other side of the cards which would make the rule true or false.

Hence, until the present experiment was conducted, only the use of concrete terms, and realistic situations, seemed to improve performance significantly in the selection task. The critical question now arises. Why does the process of discovering the rule lead to more insight than the "contrived experience" of the rule's structure used by Wason and Shapiro (1971)? An analogy may be helpful. With "contrived experience" it is as if an individual is first given the description of a maze, and then instructed to test this description by checking (or exploring) its possible and impossible paths. In our experiment, on the other hand, it is as if the individual has to generate the description of the maze from the knowledge of its three possible, and one impossible, paths. "Contrived experience" depends on a given structure, which may not be well understood, even though its components are delineated: the subject merely learns how the parts make up the whole. But in our experiment there is no whole presented initially: the subject has to construct this whole from its parts, and express it in words. One might suggest that this constructive, synthetic process allows the subject to grasp the concept of a conditional rule at a deeper level.

In particular, it seems likely that such a method prevents the subject from being seduced by superficial characteristics of the rule, and hence enables him to avoid the subsequent erroneous inferences which are so typical of the selection task (Wason, 1969*b*).

I wish to express my deepest gratitude to P. C. Wason for his valuable advice and for the radical revision of the manuscript. I am also indebted to Maria Sonino Legrenzi for help in carrying out all the experiments.

References

WASON, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 273–81.

WASON, P. C. (1969*a*). Regression in reasoning? *British Journal of Psychology*, **60**, 471–80.

WASON, P. C. (1969*b*). Structural simplicity and psychological complexity. *Bulletin of the British Psychological Society*, **22**, 281–4.

WASON, P. C. and JOHNSON-LAIRD, P. N. (1972). *Structure and Content in Reasoning: a Psychological Enquiry*. London: Batsford.

WASON, P. C. and SHAPIRO, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, **23**, 63–71.

Received 29 *April* 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 423–431

VISUAL SPAN OF APPREHENSION IN PATIENTS WITH UNILATERAL CEREBRAL LESIONS

ELIZABETH K. WARRINGTON AND P. RABIN

National Hospital for Nervous Diseases, Queen Square, London

The visual span of apprehension for random letter and digit sequences, approximations to English sequences and non-symbolic line stimuli was measured in patients with unlilateral cerebral lesions. The left hemisphere group was significantly impaired relative to the right hemisphere group and a control group on all three types of visual span task. The deficit was most marked in patients with left posterior lesions. The visual span deficits were not related to other language deficits. The findings are discussed in terms of a modality-specific defect of visual short-term memory.

Introduction

One variety of perceptual disorder described in the neurological literature appears relevant to the visual span of apprehension. Simultanagnosia, the impairment of interpretation of pictures and difficulty in reading, occurs as a specific and isolated syndrome (Wolpert, 1924). Typically affected patients, though able to point out individual features of complex pictures, cannot give an adequate account of the picture as a whole; though able to read isolated letters, they can only read whole words by laboriously spelling them out letter by letter. In an analytical study of four selected patients with simultanagnosia, Kinsbourne and Warrington (1962) demonstrated a functional deficit in simultaneous form perception. Using tachistoscopic presentation, single forms were accurately reported at normal threshold exposure durations, but simultaneous presentation of two forms resulted in failure to report one of them. Since this limitation of simultaneous perception involved both visual fields and could not be accounted for by a visuo-sensory defect or by unilateral visual neglect, it is appropriate to conceptualize the cardinal feature of this syndrome as an extreme reduction in visual span of apprehension. Though no satisfactory anatomical verification was available, there was evidence of a left posterior cerebral lesion in all cases. In one further case a localized lesion in the region of the temporo-occipital junction was found at autopsy (Kinsbourne and Warrington, 1963). The anatomical correlates of the syndrome cannot be secure without quantitative data obtained from an unselected series of cases with lesions in each cerebral sector.

More recently, the selective impairment of auditory verbal short-term memory has been described in a single patient with a left parietal lesion, (Warrington and Shallice, 1969). In this patient visual span of apprehension, though below normal, was superior to auditory span, whereas in normal subjects the reverse is the case.

Furthermore, on a short-term forgetting task there was a greater decrement over increasing intervals with auditory than with visual presentation (Warrington and Shallice, 1971). These findings were interpreted as evidence for a separate visual short-term memory system, retrieval from which does not necessarily involve the auditory short-term memory system. It is therefore important to consider whether specific *visual* short-term memory defects also occur.

Kinsbourne and Warrington (1962) demonstrated in their four patients with simultanagnosia that the visual span of apprehension was reduced for a variety of test stimuli (e.g. letters, digits, pictures of objects and geometrical shapes). These were all "verbal" stimuli in so far as all could be readily named. Hence in the present study, in addition to digit and letter stimuli, non-symbolic line stimuli were included as a relatively "non-verbal" task. Letter strings with variable transitional probabilities which were not studied previously were also included.

The present study of visual span of apprehension in patients with unilateral cerebral lesions was undertaken with three main aims:

(1) to confirm the association of defects of visual span of apprehension with unilateral lesions of the left posterior cortex;

(2) to assess the relationship between visual span of apprehension and auditory span;

(3) to consider the visual short-term memory hypothesis.

Method

Patient groups

Subjects were selected from patients referred to the Psychology Department of The National Hospital, Queen Square. Every patient with good evidence (obtained radiologically or at operation) of a unilateral cerebral lesion, able to co-operate in the test situation, and right-handed for writing was included in the series.

Thirty-nine left-hemisphere cases and 27 right-hemisphere cases, selected in this way and seen consecutively, were tested. All patients were classified by a radiologist as having a frontal, temporal, parietal or occipital lobe lesion or a combination of these four categories. The right and left hemisphere cases were further subdivided into three groups according to site of lesion, as follows:

(a) *anterior group*: cases classified as frontal, fronto-parietal or fronto-temporal.

(b) *temporal group*: cases with lesions restricted to the temporal lobe.

(c) *posterior group*: cases classified as occipital, parietal, temporo-parietal or occipito-parietal.

The left-hemisphere group included 20 cases in whom there was clinical evidence of dysphasia. Seven right-hemisphere cases and seven left-hemisphere cases had visual field defects. Most patients in both groups had neoplasms.

Twenty patients with extra-cerebral neurological disease were tested as controls.

Procedure

(a) *Visual span tasks*

A measure of visual span using tachistoscopic presentation was determined for numerical, alphabetical and simple (non-symbolic) line test stimuli. For all types of test material, strings of *five* items horizontally arranged were simultaneously exposed in the "central" field of vision.

Test 1. Random digits. Eighteen strings of five random digits (1–9) were prepared with the constraint that all digits were used equally often.

Test 2. Random letters. Fifteen strings of five random letters were prepared with the constraint that all letters (except I and O which were omitted) were used equally often.

Tests 3 *and* 4. Approximations to English. Fifteen strings of five letters were derived from Miller, Bruner and Postman (1954); the transitional probabilities of the letters were second order and fourth order approximations to English, respectively.

Test 5. Lines. Sixteen strings of five lines (straight or curved) were randomly drawn from the following set of four alternatives: —/(∪

An electronic tachistoscope with an exposure field subtending 14° × 9° of visual angle with a faint central fixation cross was used binocularly by subjects in a light-adapted state. The pre- and post-exposure fields were constantly illuminated (2·5 ft lamberts). A constant exposure duration of 50 msec for the symbolic test material (digits and letters) and 160 msec for the line stimuli was used for all subjects. The five-item strings, horizontally arranged, were presented simultaneously so that the third, i.e. middle item, coincided with the central fixation cross on the pre-exposure field.

The subject fixated the central fixation cross before each exposure and was given a "ready" signal. He was instructed to report the items in order from left to right; responses were spoken in the case of digits and letters; with the line stimuli responses were drawn on pages ruled with five columns representing the five positions of the items in the sequence.

(b) *Verbal tasks*

(1) *Digit span*. The digit span sub-test of the W.A.I.S. (Wechsler, 1955) was given and the number of digits (forward recall) was recorded as the measure of auditory span.

(2) *Vocabulary*. The vocabulary sub-test of the W.A.I.S. was given and the weighted score (corrected for age) was recorded as a measure of expressive speech function.

(3) *Naming*. A set of 20 line drawings from the Oldfield and Wingfield (1965) series was used as a test of recognition and naming. Each response was scored as being either correct, an error of naming, or an error of recognition. A response was recorded as an error of naming if the subject was able to indicate, by description or gesture, that he knew what the picture represented though unable to name it.

Results

Mean scores on the three verbal tasks, digit span, vocabulary and naming pictures, for each patient group, are given in Table I. The difference between the left-hemisphere group and right-hemisphere group is significant at the 1, 5 and 0·1% levels respectively (Mann-Whitney U-Test) on digit span, vocabulary and naming.

TABLE I

Mean scores on verbal tasks

	Vocabulary	Digit span	Naming errors
L	9·4	5·8	3·5
R	11·6	7·1	0·4
C	11·6	7·0	0·1

The control group and right-hemisphere group obtain almost identical scores on these tasks. On the digit span task there are no significant differences according to site of lesion within the left hemisphere, though on the vocabulary sub-test the left temporal group was significantly impaired relative to both the left posterior group and the left anterior group ($P<0{\cdot}05$).

The percentage correct (items in correct serial position) for each visual span test was recorded. In the lines test only, a guessing correction was applied.* The mean percentage correct for each patient group and subgroup for each visual span test are given in Table II. The Mann-Whitney U-Test was used in all group comparisons and values of z and their probability are given in Table III. The left-hemisphere group performed significantly worse than the right-hemisphere group not only on the four tests using numerical and alphabetical material test stimuli but also on the lines test. On no visual span test, including the lines test, was the right-hemisphere group significantly worse than the control group.

TABLE II

Visual span: mean % correct in patient groups

Groups / Stimuli	Random digits	Random letters	2nd-order letters	4th-order letters	Lines
Control $N = 20$	80	69	90	92	56
Right $N = 27$	74	68	88	92	55
Left $N = 39$	55	50	73	77	43
R. Anterior $N = 8$	81	78	88	93	53
L. Anterior $N = 16$	63	56	80	87	49
R. Temporal $N = 12$	74	66	91	94	61
L. Temporal $N = 10$	54	50	78	73	44
R. Posterior $N = 7$	65	65	85	88	49
L. Posterior $N = 13$	44	41	62	62	35

* (Number correct − number of false positives)/5, see Guilford (1936).

TABLE III

Comparison of patient groups on visual span tasks

	Random digits	Random letters	2nd-order letters	4th-order letters	Lines
L. vs R.	$z = 3.08$†	$z = 3.98$‡	$z = 3.46$‡	$z = 2.72$†	$z = 2.79$†
L. vs C.	$z = 4.17$‡	$z = 3.58$‡	$z = 3.75$‡	$z = 2.51$†	$z = 3.31$‡
R. vs C.	$z = 1.3$	$z = 0.94$	$z = 1.44$	$z = 0.86$	$z = 0.27$
R. Posterior vs. L. Posterior	$z = 2.50$*	$z = 2.62$*	$z = 2.16$*	$z = 1.87$	$z = 2.10$*
R. Temporal vs. L. Temporal	$z = 2.50$*	$z = 1.91$	$z = 2.83$†	$z = 2.80$†	$z = 1.90$
R. Anterior vs. L. Anterior	$z = 2.02$*	$z = 1.77$	$z = 1.04$	$z = 0.67$	$z = 0.81$
R. Posterior vs. R. Temporal and R. Anterior	$z = 1.61$	$z = 1.63$	$z = 0.86$	$z = 0.94$	$z = 1.04$
L. Posterior vs. L. Temporal and L. Anterior	$z = 2.26$*	$z = 2.41$*	$z = 2.04$*	$z = 2.35$*	$z = 2.26$*

* $P < 0.05$
† $P < 0.01$
‡ $P < 0.001$

Though both the left temporal and the left posterior groups were consistently impaired relative to the right temporal and right posterior groups respectively, the left posterior group was significantly worse than the left non-posterior group, indicating a focal deficit in the left posterior group. The relevant control comparison of the right posterior group with the right non-posterior group was not significant.

Correlations between visual span and language tests

The median test of significance was used to assess the degree of association between performance on the visual span task (random sequences) and dysphasia (apparent clinically) and the naming task. Neither association was significant.

The degree of association between performance on the three types of visual span task [(1) random sequences: the combined score on random digits and random letters; (2) sequences approximating to English: the combined score on second-order and fourth-order letters; (3) sequences of lines] and the vocabulary score and the auditory digit span score was measured using the product moment correlation. The values of r for each patient group are given in Table IV.

The correlations between the three visual span measures and the auditory digit span and vocabulary are generally low, and in many cases there is not a significant association. The correlation between vocabulary and the visual span for random

TABLE IV

Correlations between visual span tasks and verbal tasks: auditory digit span and vocabulary

	Auditory digit span			Vocabulary		
	Random sequences	Approximations to English sequences	Line sequences	Random sequences	Approximations to English sequences	Line sequences
Left hemisphere $N = 39$	0·36*	0·11	0·25	0·15	0·33*	−0·08
Right hemisphere $N = 27$	0·39*	0·48*	0·15	0·32	0·12	−0·02
Controls $N = 20$	0·61†	0·35	0·31	0·45*	0·32	0·46*

* $P < 0{\cdot}05$
† $P < 0{\cdot}01$

sequences measures is somewhat lower in the left-hemisphere group than in the control group, though the left-hemisphere group as a whole was impaired on both types of task. There are no significant differences between the three patient groups in the correlations between auditory span and visual span. However, in the left-hemisphere group, who were impaired on both tasks, there is a trend for the degree of correlation in the left-hemisphere group to be lower than in the control group. In the left-hemisphere group only one correlation (auditory digit span with visual span for random sequences) reached significance, and in the critical left posterior group ($N=13$), in whom the deficit on visual span tasks was most marked, the correlation between auditory span and visual span for random sequences was 0·39, which is not significant ($P>0{\cdot}2$).

TABLE V

Correlations between visual span for random sequences and other span tasks

	Random sequences vs. lines	Random sequences vs. approximations to English
L	$r = 0{\cdot}75$‡	$r = 0{\cdot}72$‡
R	$r = 0{\cdot}63$†	$r = 0{\cdot}75$‡
C	$r = 0{\cdot}50$*	$r = 0{\cdot}77$‡

* $P < 0{\cdot}05$
† $P < 0{\cdot}01$
‡ $P < 0{\cdot}001$

Correlations between visual span tasks

The correlations (product moment) between performance on the random sequences and both English approximations and line sequences were significant in all patient groups (see Table V). The degree of association between the three visual tasks is clearly higher, especially in the left-hemisphere group, than between the auditory span task and any of the measures of visual span.

Visual field defects

There was a clinically detectable visual field defect in seven cases of both the left and right hemisphere groups. In neither group were visual field defects significantly associated with performance on the visual span task ($\chi^2 = 3{\cdot}4$ and $0{\cdot}6$, respectively, $df = 1$).

Discussion

Impairment of visual span of apprehension has been found in patients with left posterior cerebral lesions. All three types of visual span tasks were significantly impaired in the left posterior sub-group compared with the right posterior and the left non-posterior sub-groups. This finding confirms the earlier clinical reports of an association between limited visual span and lesions of the left posterior cerebral region (Kinsbourne and Warrington, 1962).

It can be shown that visual span of apprehension is not secondary to other concomitant disabilities. The present findings cannot be attributed to visual field defects *per se* in so far as there were roughly equivalent numbers of patients with field defects in the two groups and there was no association between performance on the visual span tasks and the presence of a visual field defect. Though many of the patients in the left-hemisphere group were moderately aphasic, this disability as assessed by a naming task was not associated with impaired visual span. On the vocabulary test the left posterior sub-group were not specifically impaired, whereas the left temporal group were. The converse was found on the visual span task. What then is the basis of the reduction of visual span of apprehension? Three alternative possibilities are considered.

On tasks of visual perception with no memory component patients with right-hemisphere lesions are impaired compared with patients with left-hemisphere lesions. Patients drawn from the same population and according to identical criteria as the present series have been assessed on threshold measurements for the visual detection of a dot, and for single letter recognition, and no consistent deficit associated with localization of lesion was found (Warrington and Rabin, 1970). However, on a dot subiting task, in which only the number of test stimuli must be reported, there was no deficit in the left parietal group (which is most comparable to the present left posterior group) whereas the right parietal group was significantly impaired (Warrington and James, 1967). The possibility that impaired visual-sensory functions could account for the present test findings is therefore rejected.

The question of whether an agnosia for letters (adequate perception but a failure of identification) can account for the present findings is more difficult to answer. In the original clinical studies of simultanagnosia (Kinsbourne and Warrington,

1962) emphasis was placed on the discrepancy between single and double form perception, including letters, no defect of single letter recognition being demonstrated. Nevertheless, the possibility remains that a very mild letter identification defect results in a limitation in span. The plausibility of this suggestion is weakened both by the original observations that the disorder of simultaneous form perception is not specific to particular types of stimuli, which is the case with agnosic identification defects (Hécaen, Ajuriaguerra and David, 1952). The present findings, that performance on non-symbolic line stimuli and symbolic stimuli is equally impaired in the left hemisphere group, and that performance on the span for line sequences correlates with performance on random sequences, would be difficult to account for in terms of an agnosic defect. It is therefore argued that it is unlikely that a minor form of a letter identification impairment is the basic defect on the visual span task.

It is suggested that a short-term visual memory deficit can account for the reduced visual span of apprehension. One of the cardinal features of short-term memory is a limitation in capacity, related to the nature of the test stimuli. Using simultaneous presentation of an array of stimuli it can be shown that not only can subjects "store" more than they can report (Sperling, 1960), but the same variables affect performance of visual short-term memory tasks as of visual span of apprehension. Indeed, Baddeley (1964) argues that the increase in span of apprehension found with increasing approximation to English is due to a decrease in the "memory load" of the task. One critical difference between the visual span task and other perceptual tasks hitherto investigated in patients with cerebral lesions is the "memory" component of the task. The finding that the variables which decrease span (non-symbolic line stimuli) and increase span (approximation to English) in normal subjects also operate in the left posterior group, which showed a specific impairment in visual span, is more compatible with an interpretation in terms of a visual memory defect than a perceptual impairment or defective letter identification.

Sperling (1967) has argued on the basis of tachistoscopic visual span of apprehension experiments that it is necessary to postulate at least three short-term memory stores contributing to this task: namely, the iconic store, a recognition buffer store, and an auditory verbal store. Which of these stores could be implicated in the present findings?

It is not plausible to suggest that a defect of auditory short-term memory could account for the present findings. The left posterior hemisphere group was significantly impaired on the visual span task but not on the auditory span measure and furthermore the correlation between auditory and visual span was only significant for the random sequences (and in the most impaired sub-group of left posterior cases it was not significant). These findings indicate a lack of interdependence of auditory and visual short-term memory functions. It is unlikely that the iconic store is implicated in the present impairment of visual span since it was shown in earlier studies (Warrington and James, 1967) that dot subiting (enumerating the number of stimuli in an array) was not specifically impaired in the left parietal cases, and indeed, in the orignal clinical studies patients with a visual span of one were able to dot subit relatively normally (Kinsbourne and Warrington, 1962). Sperling's (1967) notion that a third store is required to account for visual

span data is supported by recent findings in the patient K.F., in whom there was evidence of a store comparable, at least in duration, to the auditory store (Warrington and Shallice, 1971). This was shown to have the characteristics of a "visual store" rather than a store of "motor programme instructions" as suggested by Sperling. It is suggested that the most plausible explanation for the present findings is impairment of a visual short-term memory system. That visual store, shown to be intact and independent of auditory S.T.M. in K.F., may be impaired in the present left posterior group. This hypothesis could be put to a more direct test by employing recognition, partial report or visual search techniques.

We wish to thank the physicians and surgeons of the National Hospital for permission to study and report our findings on patients under their care. We are grateful to Dr R. T. C. Pratt and Dr T. Shallice for helpful advice in the preparation of this manuscript. We thank Mrs A. M. Taylor and Miss V. Resin for their assistance with the statistical analysis. The late Mrs Pauline Rabin was supported by a grant from the Medical Research Council.

References

BADDELEY, A. D. (1964). Immediate memory and the perception of letter sequences. *Quarterley Journal of Experimental Psychology*, **16**, 364–8.

GUILFORD, J. P. (1936). The determination of item difficulty when chance success is a factor. *Psychometrika*, **1**, 259–64.

HECAEN, J., AJURIAGUERRA, J. de and DAVID, M. (1952). Les deficits fonctionels apres lobectomie occipitale. *Monatsschrift für Psychiatrie und Neurologie*, **123**, 239–91.

KINSBOURNE, M. and WARRINGTON, E. K. (1962). A disorder of simultaneous form perception. *Brain*, **85**, 461–486.

KINSBOURNE, M. and WARRINGTON, E. K. (1963). The localizing significance of limited simultaneous visual form perception. *Brain*, **86**, 697–702.

MILLER, G. A., BRUNER, J. S. and POSTMAN, L. (1954). Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology*, **50**, 129–39.

OLDFIELD, R. C. and WINGFIELD, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, **17**, 273–81.

SPERLING, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, **74**, no. 11.

SPERLING, G. (1967). Successive approximations to a model for short-term memory. *Acta psychologica*, **27**, 285–92.

WARRINGTON, E. K. and JAMES, M. (1967). Tachistoscopic number estimation in patients with unilateral cerebral lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, **30**, 468–74.

WARRINGTON, E. K., KINSBOURNE, M. and JAMES, M. (1966). Uncertainty and transitional probability in the span of apprehension. *British Journal of Psychology*, **57**, 7–16.

WARRINGTON, E. K. and RABIN, P. (1970). Perceptual matching in patients with unilateral cerebral lesions. *Neuropsychologia*, **8**, 475–87.

WARRINGTON, E. K. and SHALLICE, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain*, **92**, 885–96.

WARRINGTON, E. K. and SHALLICE, T. (1972). Neuropsychological evidence of visual storage in short-term memory tasks. *Quarterly Journal of Experimental Psychology* (In press).

WECHSLER, D. (1955). Wechsler Adult Intelligence Scale. New York: The Psychological Corporation.

WOLPERT, I. (1924). Die Simultanagnosie-Storung der Gesamtauffassung. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, **93**, 397–415.

Received 10 June 1971

naires relating to events of each two-year period sampled was recorded. The mean percentage correct for each age group for each questionnaire is given in Figure 1. The effects of passage of time since the event and of age of the subject are clear cut. Performance declines fairly rapidly for events of the preceding eight years and then remains relatively stable except for events of the war years for which recall is rather better. This better recall of events of the war years is also found in the school children who did not experience those events. Except for the two questionnaires (1967–68 and 1965–66) relating to events they were old enough to

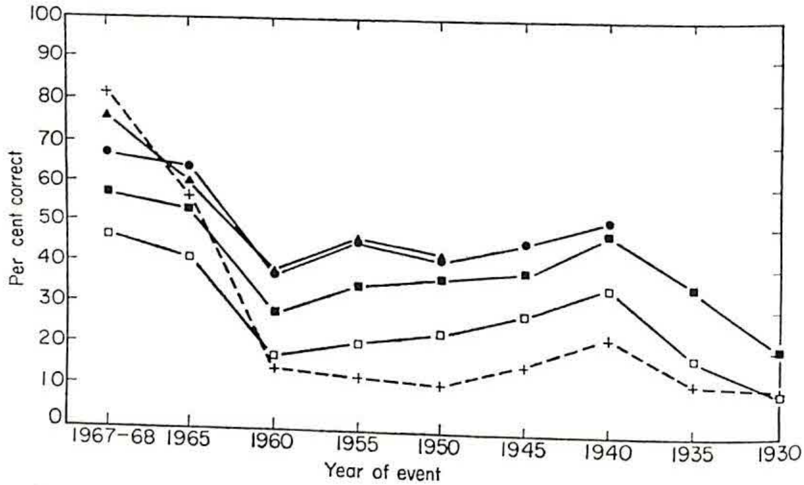

FIGURE 1. Long term memory questionaire—multiple choice ○ 40s ● 50s ■ 60s □ 70s ▲ Average 16 yrs 8 mths.

experience, the children's mean scores were worse than any adult subject groups.

Performance on this task is directly related to age for all questionnaires. The forgetting curves for each age group are approximately in parallel. The impairment in performance with increasing age of subject found for recent events (i.e. 1967–68) does not decrease or change as the events sampled become more remote. In particular there is no indication that the oldest age group, 70–80, are relatively less impaired on questionnaires relating to any of the more remote years sampled than on the questionnaires concerning the most recent events.

A series of *t*-tests were computed comparing adjacent age groups for each of the nine questionnaires (see Table I). There are no significant differences between the 40 and 50 age groups at any point in time. The 60-year olds are significantly worse than the 50-year olds on all questionnaires except for the most remote events tested (1940–41). However, the 70-year olds are significantly worse than the 60-year olds on all questionnaires including the most remote. The degree of deficit, as indicated by the level of significance of the differences, is relatively similar for all points in time.

It was noted that a subject's performance on past years could be predicted by his performance on the questionnaire for the most recent events (1967–68). Correlations were therefore computed between scores on the most recent questionnaire and the earlier questionnaires for each age group (Table II). There is a highly

TABLE I

L.T.M.Q.—Comparison of adjacent age groups—values of t

Age	Year	1967–68	1965–66	1960–61	1955–56	1950–51	1945–46	1940–41	1935–36	1930–31
70's v 60's	Recall	3·3†	3·48‡	2·96†	3·58‡	2·23*	2·69†	3·05†	2·05*	4·35‡
	Recognition	2·2*	2·98†	2·7*	1·61	3·23†	3·5‡	1·45	1·0	2·1
60's v 50's	Recall	3·38‡	3·94‡	3·13†	3·87‡	2·50*	2·38*	1·1		
	Recognition	4·83‡	4·41‡	3·18†	4·31‡	2·37*	2·75†	2·59†		
50's v 40's	Recall	2·43*	1·14	0·86	0·16	0·45				
	Recognition	0·16	1·78	1·04	0·91	0·10				

* Significance at 5%; † significance at 1%; ‡ significance at 0·1%.

significant correlation in all age groups between memory for recent events and all more remote events.

Recognition

The mean percentage correct on the multiple choice version of the questionnaire for each adult age group and the school children is shown in Figure 2. Comparing events of the recent past (1967–68) with those of 1965–66 and 1960–61 there is a decrement in performance in all age groups, but little further decrement with increasing time intervals. On the contrary retention for events relating to the war years is rather better. The effect of age is not so great as for the recall version of the questionnaire. Nevertheless significant differences between the 50 and 60

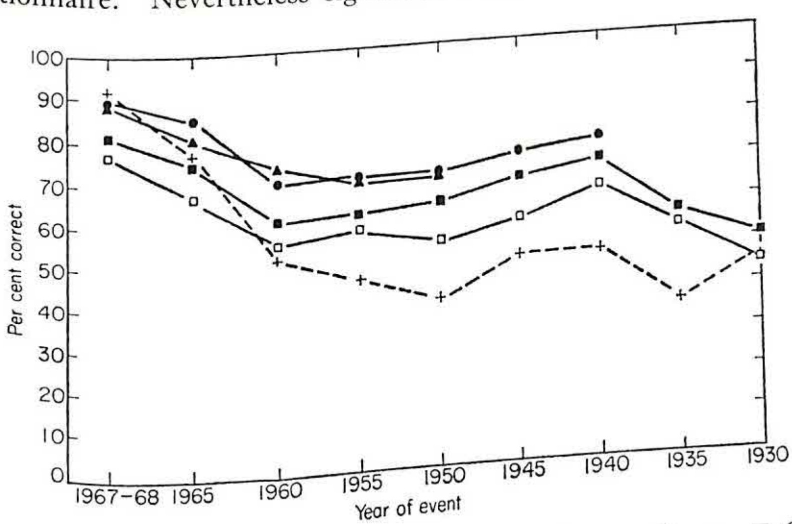

FIGURE 2. Long term memory questionnaire—free recall. ○ 40s ● 50s ■ 60s □ 70s ▲ Average 16 yrs 8 mths.

age groups were found for all time intervals, and for all but three sets between the 60 and 70 age groups. Performance on the most recent set was significantly correlated with all the more remote sets in all age groups (Table II).

TABLE II

L.T.M.Q.—Correlation of performance on contemporary events (1967-68) *with more remote events—values of* r

Age	Year	1965 –66	1960 –61	1955 –56	1950 –51	1945 –46	1940 –41	1935 –36	1930 –31
70's $N =$	Recall	0·77‡	0·65‡	0·79‡	0·74‡	0·73‡	0·72‡	0·75‡	0·57‡
50	Recognition	0·61‡	0·35†	0·48‡	0·46†	0·43†	0·57‡	0·55‡	0·44‡
60's $N =$	Recall	0·66‡	0·69‡	0·79‡	0·75‡	0·68‡	0·58‡	0·55‡	0·44‡
100	Recognition	0·41‡	0·38‡	0·72‡	0·39‡	0·36‡	0·26‡	0·42‡	0·41‡
50's $N =$	Recall	0·69‡	0·60‡	0·63‡	0·66‡	0·59‡	0·56‡		
100	Recognition	0·51‡	0·47‡	0·31†	0·45‡	0·58‡	0·40†		
40's $N =$	Recall	0·63‡	0·61‡	0·59‡	0·54‡				
54	Recognition	0·41†	0·26	0·41†	0·71‡				

† Significance at 1%; ‡ significance at 0·1%.

Memory for Well-known Faces

Method

The principle applied in the design of the questionnaire for past events has been adapted to examine recent and remote visual memories. Recognition of contemporary personalities can be compared with photographs of persons no longer in public view. By choosing personalities whose public life was sharply terminated at a finite time in the past, one can achieve a test of recognition of faces at varying intervals of time since they were learned or experienced. Apart from the intrinsic interest of investigating visual memory it was hoped that such material would be less subject to repeated review in the mass media and in books, and hence the forgetting functions might be less contaminated by recent rehearsal. Again school children have been tested in order to differentiate the "memory" and "history" components of the task.

Design of faces recognition test

A pilot study was carried out to determine which contemporary personalities were reliably known from their photographs. 15 faces frequently recognized were selected to represent as varied occupations and roles as possible. Sets of 15 personalities who had disappeared from public view in the early 1960s, the late 1950s, the early 1950s and the late 1940s respectively were chosen. Thus the test comprised 75 photographs of well-known personalities, representing 5 time intervals since they were experienced.

As with the L.T.M.Q. there were two versions of the test, identification (recall) and identification by multi-choice (recognition). In the multi-choice version 3 alternative names, 1 correct and 2 incorrect, were presented auditorially, while the subject was inspecting the photograph. The incorrect alternatives were chosen to be both plausible and equally well-known by name as the correct name.

Subjects

(1) Two hundred adult subjects between the ages of 40 and 89 were tested and subdivided into four groups of 50 subjects according to decade of life, except for the oldest age group which comprised 50 subjects between the ages of 70 and 89. The same source of subjects who completed the L.T.M.Q. was used for the Faces Recognition test. (2) Fifty boys and girls in the 1st year of a 6th form at a grammar school were tested. All subjects (mean age 16 years 11 months) were doing "A" level courses.

Procedure

Adult subjects were tested individually, between January 1970 and July 1970, the experimenter recording the responses. For each photograph recall by identification was attempted first, followed immediately by the multi-choice. Three alternative names in random order were presented for all photographs not correctly identified, 1 correct, 2 incorrect, the subject being required to respond to each choice. Faces correctly identified were assumed to be correct on the multi-choice version. The school children were tested (October 1970) on both versions of the face recognition test in a group. First the photographs were projected on to a screen and 15–30 sec were allowed for photographs to be identified (name written on record sheet). Faces were presented in a random order, the contemporary faces alternating with personalities from the past. The series of faces were then presented a second time, the 3 alternatives being presented auditorially while the photograph was inspected. The identification by choice was again written on a record sheet.

Results

Recall

The number of faces correctly identified, expressed as a percentage, for each age group is shown in Figure 3. The effect of time interval is clear cut, a marked decrement occurring between performance on contemporary faces and faces of the early 1960s, and still further decrement with the more remote faces. Performance is clearly related to age of subject, each decade of life resulting in some further impairment of performance, the effect being most marked in the oldest age group (70–89). Furthermore the forgetting functions are very similar in the adult

TABLE III

Memory for faces—Comparison of adjacent age groups—values of t

		Contemporary	Early 60's	Late 50's	Early 50's	Late 40's
70 v 60	Recall	5·56‡	4·48‡	5·35‡	4·75‡	4·92‡
	Recognition	2·30*	4·8‡	3·38†	4·22‡	3·83‡
60 v 50	Recall	0·93	0·35	1·3	1·42	1·2
	Recognition	1·0	1·7	2·4	1·4	0·98
50 v 40	Recall	1·3	2·07*	1·3	0·73	0·48
	Recognition	0·57	0·26	0·26	0·25	0·76

* Significance at 5%; † significance at 1%; ‡ significance at 0·1%.

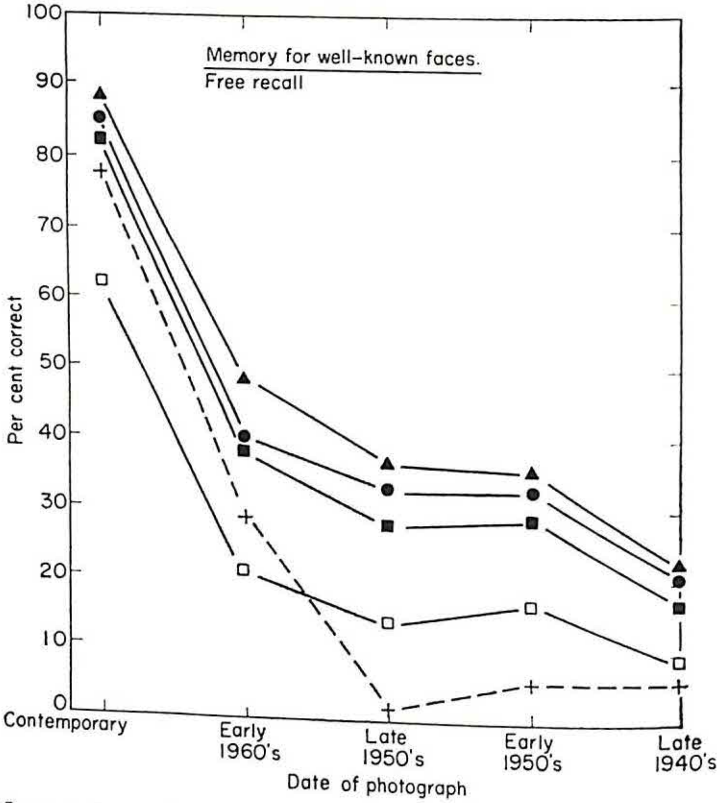

FIGURE 3. Memory for well-known faces. Free recall. ▲ 40s ● 50s ■ 60s □ 70s-80s + Average 16 yrs. 11 mths.

TABLE IV

Memory for faces—Correlation of performance on contemporary faces with more remote faces—values of r

Age	Year	Early 60's	Late 50's	Early 50's	Late 40's
70's $N =$ 50	Recall	0·59‡	0·39†	0·45‡	0·48‡
	Recognition	0·31*	0·28*	0·20	0·26
60's $N =$ 50	Recall	0·68‡	0·51‡	0·68‡	0·46‡
	Recognition	0·49‡	0·39†	0·45‡	0·34*
50's $N =$ 50	Recall	0·41†	0·55‡	0·13	0·41‡
	Recognition	0·75‡	0·59†	0·71‡	0·44‡
40's $N =$ 50	Recall	0·59‡	0·51‡	0·49	0·46‡
	Recognition	0·90‡	0·90‡	0·69‡	0·68‡

* Significance at 5%; † significance at 1%; ‡ significance at 0·1%.

group, the four curves being approximately parallel. The oldest age group obtain the lowest mean score not only on contemporary faces but on all the earlier time periods. Performance of adjacent age groups (40 and 50 years) are significantly different for only one time period (the early 1960s) and the 50 and 60 age groups do not differ significantly for any time period. However, there is a highly significant difference between the 60 age group and the 70+ age group on all time periods. Performance on the contemporary faces was correlated with performance on the four earlier sets of faces. In all age groups (except the 50-year olds on faces of the early 1950s) there was a highly significant correlation between performance on the contemporary faces and on the more remote sets of faces (see Table IV). The degree of correlation is much the same for the older as for the younger subjects.

The performance of the school children for the three earliest sets of photographs is less good than that of all adult groups, suggesting that remote faces are not as subject to recapitulation as may be the case for events.

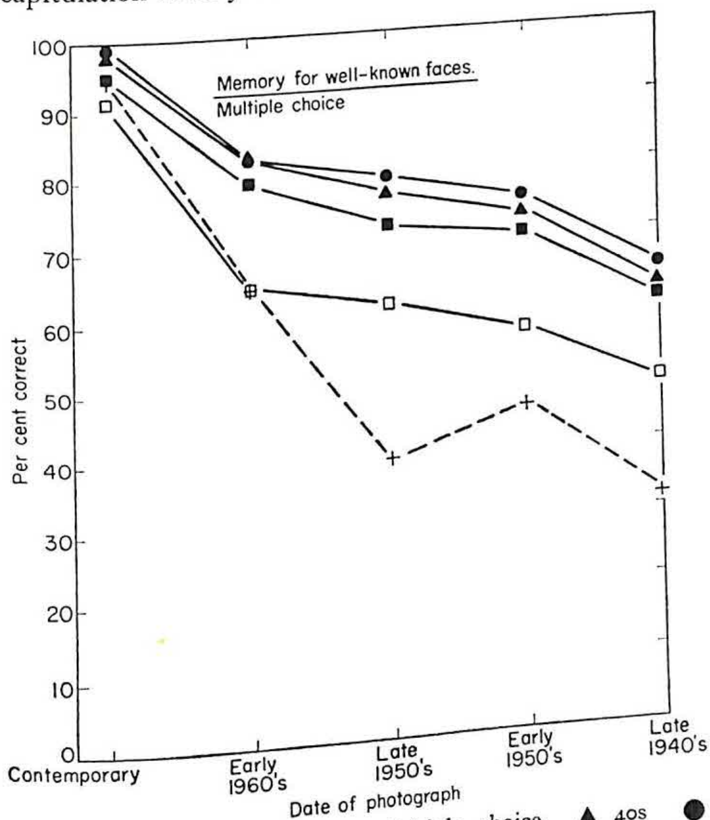

FIGURE 4. Memory for well-known faces. Multiple choice. ▲ 40s ● 50s ■ 60s □ 70s–80s + Average 16 yrs. 11 mths.

Recognition

The number of faces correctly identified, expressed as a percentage, on the multi-choice version of the task for each age group is shown in Figure 4. Performance for all sets of faces is clearly superior to recall. There is nevertheless a substantial decrement related to the remoteness of the experience. Performance

on the contemporary faces, except in the oldest age group, is above 95% correct, yet on the late 1940s set the mean % correct has dropped to 50% in the oldest age group, and between 65 and 70% in the other age groups. The slight difference in performance between the 40 and 50 age groups is not significant, and the 50's and 60's differ significantly on only one set of faces (late 1950s), whereas there is a highly significant difference between the 60 and 70 age groups on all sets except the contemporary faces. This is only just significant at the 5% level (see Table III). The correlation between performance on the contemporary set of faces and the earlier sets was significant in all age groups except the 70s, who were relatively good on the contemporary set but showed a clear impairment on the more distant faces.

Discussion

Memory for recent and remote events has been assessed in normal subjects ranging in age from 40 to 80, using questionnaire techniques. No evidence for the preservation of remote events in the older age group has been obtained. The same result was obtained using a very different task, recognition of well-known faces; the older subjects were clearly impaired in recognizing faces from the more distant past. Ribot's law, that the dissolution of memory is inversely related to the recency of the event, requires only that remote events be relatively preserved and recent events relatively more vulnerable. However, the present findings indicate that not only are the older subjects impaired on tasks relating to all time periods sampled, but there was not even a trend for this deficit to be less marked for the more remote events. Similarly on the well-known faces task there was no indication of greater impairment on contemporary faces compared with faces from the more distant past.

This study has a number of limitations. First, it is assumed that the nine sets of questions relating to public events were of equivalent difficulty. An inadvertent increase in difficulty with increase in time interval since the event could account for the present findings. Such an experimental bias in a retrospective study seems impossible to assess. However, with the memory for well-known faces task it was possible to approximate to matched "sets", for which the age and passage of time effects were at least as convincing as for memory for events. Second, one cannot exclude the possibility that had memory for personal events been assessed a different pattern of results would have emerged. Public events were chosen in preference to a technique based on personal events as they provide a common source of experience for all subjects. It seems unlikely that memory for such experience should be subject to different laws than experiences unique to the individual. Third, though the total time period sampled in this study is relatively long (25 years in the well-known faces task and 38 years in memory questionnaire) memory for even more remote events might be less vulnerable in the older age group. Certain practical problems arise in extending the questionnaire to more remote events. One cannot reasonably compare events experienced as an adult (the 70-year olds in the 1920's) with events experienced as a child (the 60-year olds in the 1920's). Some alternative techniques are required to assess memory for time periods extending beyond 40 years. In common with all studies of memory and ageing,

there is the impossibility of equating the three factors, age at learning, retention interval and age at testing.

The question arises as to the validity of these techniques. The assumption is made that memory for past experiences is being tested rather than knowledge which could have been recently acquired. Does the "memory" component outweigh the "historical knowledge" component? School children who were not old enough to have experienced any but the most recent events and the contemporary faces were therefore tested to assess the opportunity for recent acquisition. In general the school children were comparable with the adults for contemporary events and contemporary faces, but for more remote events their "memory" was less good. This finding was particularly clear cut for the well-known faces task, where there was negligible recall of personalities of the 1940's and 1950's. It is therefore argued that performance on these tasks, particularly the faces, is largely determined by memory for past experiences, and that opportunities for recent acquisition are relatively unimportant.

These findings have implications for theories regarding memory processes and ageing. On the basis of differential effects obtained using recall and recognition procedures, it has been suggested that the memory impairment in older subjects is one of retrieval (Craik, 1971; Schonfield and Robertson, 1966). In a previous study using similar questionnaire techniques (Warrington and Silberstein, 1970) it was found that on the recall version the older subjects were clearly worse than the younger subjects, but not on the multi-choice version. However, in the present investigation the differential effects of recall and multi-choice tasks were less striking; there was a significant impairment on the multi-choice version of the questionnaire in the 60-year olds and on the multi-choice version of the well-known faces test in the 70-year olds. These findings suggest that the poor memory of older subjects is not entirely a retrieval difficulty, though it may indeed be prominent.

The present findings, although not relevant to theories of forgetting, go some way towards eliminating the problem of accounting for the relative preservation of remote memories. It is suggested that in the dissolution of memory a unitary process may be implicated. Performance on new learning tasks is correlated with performance on remote memory tasks. That is, the efficiency of memory for the most recent events can predict the efficiency of memory for remote events in all age groups.

The sparing of remote memories has been claimed not only in normal old people, but also in amnesic subjects. Retrograde amnesia, the loss of memory for time periods antedating the illness, has been reported to be of some years' duration, with preservation of more remote experiences. Five severely amnesic subjects without any generalized intellectual impairment were tested on both the questionnaire for events and the well-known faces test, and it was found that the amnesia extended for the whole time period sampled by these tasks. No evidence of relative sparing of the most remote experiences could be detected (Sanders and Warrington, 1971). Thus in two populations sampled, normal old people and amnesic patients, it has been shown that memory for remote events is no less vulnerable than recent events. The present findings, namely that remote memory is not differentially spared in older subjects, therefore has some generality.

We wish to thank Dr R. L. Evans, Headmaster of Great Manor Grammar School, Hanwell, the Medical Committee of the Royal National Orthopaedic Hospitals, and the Welfare Department of the London Borough of Camden for providing facilities for testing subjects. We are grateful to Dr A. R. Jonckheere for statistical advice, and to Drs F. I. M. Craik and R. T. C. Pratt for their comments on the manuscript. The Medical Research Council provided financial support.

References

Craik, F. I. M. (1971). Age difference in recognition memory. *Quarterly Journal of Experimental Psychology*, **23**, 316–23.

Ribot, T. A. (1882). *The Diseases of Memory*. New York: Appleton.

Sanders, H. I. and Warrington, E. K. (1971). Memory for remote events in amnesic patients. *Brain*. (In press.)

Schonfield, D. (1969). In search of early memories. *VIIIth International Congress of Gerontology*. Washington D.C.

Schonfield, D. and Robertson, B. (1966). Memory storage and ageing. *Canadian Journal of Psychology*, **20**, 228–36.

Warrington, E. K. and Silberstein, M. (1970). A questionnaire technique for investigating very long term memory. *Quarterly Journal of Experimental Psychology*, **22**, 508–12.

Received 23 April 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 443–448

HEMISPHERE FUNCTION AND VIGILANCE

STUART DIMOND AND GRAHAM BEAUMONT
University College, Cardiff

A vigilance task in which successive signals were presented to one or other hemiretina, and therefore to one or other cerebral hemisphere, revealed no differences between the hemispheres in terms of detections, although detections declined overall during the experimental period. False positive responses also declined, but consistently more arose from the left hemisphere. There was also a difference in the detection of signals received through the nasal and temporal hemiretinae, the temporal hemiretina showing superiority in detection rate throughout the experiment. This finding may provide a new and more economical approach to the tunnel vision phenomenon.

Introduction

Levy-Agresti and Sperry (1968) have suggested on the basis of their studies of split-brain patients that the two cerebral hemispheres employ different perceptual strategies. A subject was given a task in which he was required to feel a wooden block with either his right or left hand and to specify which of three blocks drawn on a card he had been given. The right hemisphere showed proficiency on those block sets having complex qualities which could be readily visualized, whereas the left succeeded on block sets requiring careful analysis and failed on those which demanded a more immediate grasp of spatial relationships. The pattern of successes and failures in the performance of the right hemisphere resembled more closely that of the right hemisphere of other subjects than that of its own partner, the left. This led to the suggestion that the hemispheres deal differently with information reaching them from the external world. The right, with its grasp of spatial relationships, apprehends events in a "Gestalt" fashion, whereas the left carries out a sequential analytic procedure.

Differences between the hemispheres are important, but so also is the question of how they co-operate and integrate their functions in the normal brain. This question has been taken up in a series of papers which show that response times are faster when two signals are distributed between the hemispheres than when they are both projected to the same hemisphere (Dimond, 1970). Total output of the brain was increased while performing two tasks if the perceptual load of one task was divided between the hemispheres, instead of being directed to one hemisphere alone (Dimond and Beaumont, 1971). When two signals for response were projected at the same time to one hemisphere, blocks of function were observed. This was true not only of reaction time tasks, but also of verbal tasks (Dimond, 1971). These findings suggest not only that each hemisphere allocates attention independently of the other in the normal brain, in order to analyse information entering into it, but also that each proceeds some way with the analysis before sharing or communicating with the other.

One important factor, not in respect of false alarms, but with regard to the percentage of signals detected, concerns the difference in the response to signals at the nasal or temporal hemiretinae. This is illustrated in Figure 3. There was an overall difference which was significant ($F = 26{\cdot}6$, $df = 1, 3$, $P < 0{\cdot}05$). From Figure 3 it can be seen that a difference was already apparent during the first section of the test period, and that this difference increased as the test proceeded. That the results might have been an artifact of the performance of low voltage bulbs must be considered. However, for all but the 200-msec period during which the bulbs brightened, the bulbs were "run" at 2 V below their rated voltage. Under these conditions the chances of instability or differential performance over time, able to be perceived by a human subject, may be discounted.

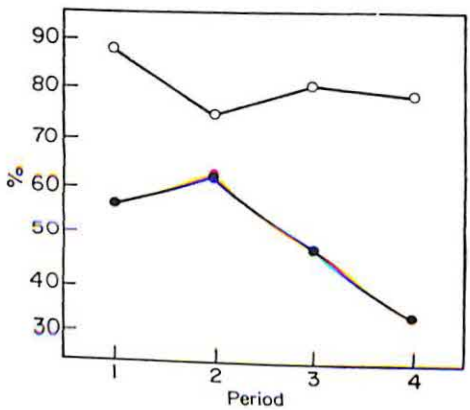

FIGURE 3. Mean percentage detection by the temporal and nasal hemiretinae, analyzed by four consecutive periods (○ = temporal hemiretina; ● = nasal hemiretina).

Discussion

While there were no differences in performance between the hemispheres in the number of signals detected, there were differences in the number of false positives to which each hemisphere gave rise, the left being responsible for more false positive responses than the right. White (1969) reported an analysis of visual hemifield responses in terms of signal detectability and failed to discover any appreciable differences in sensitivity. In the results reported here there were no significant differences in the number of signals detected by the right and left recipient hemispheres.

When differences in false positives are considered, it is commonly assumed that the false positive response is a reflection of the sensitivity of the perceptual system and of the degree to which signals can be differentiated from hypothetical neural noise. If this is accepted, then there are several explanations which may be proposed. The perceptual systems of the left hemisphere may contain more neural noise than those of the right. This would be reflected, however, by a deterioration in sensitivity and in the number of signals detected, which was not found. This fact suggests that either the right and left are equal in their capacity to detect signals, or that a common mechanism detects both right and left signals.

Another approach has been to suggest that the number of false positive responses represents a learning effect since they tend to decrease over sessions (McGrath, 1963). There are several serious objections to this view, in that situations which facilitate learning in the ordinary course of events do not appear to affect the number of false positives in quite the same way. It would be informative to trace differences in learning capacity between one hemisphere and the other, in the light of this.

A different view might be that more false positives arose because the left hemisphere bears responsibility for the initiation of motor responses. If it is keyed to trigger action, it is more motor impulsive, and on occasion it may release a response without fitting it properly to the stimulus sequence. However, irrespective of the explanatory hypothesis adopted, we are forced to conclude that the left hemisphere is not the equal of the right in this aspect of perceptual organization.

Next to consider are the differences between detections by the temporal and nasal hemiretinae. In previous experiments, measuring reaction time and employing stimuli presented singly to each hemiretina, or words presented to only one hemiretina, no significant differences among the hemiretinae were found. However, there are some reports in the literature of such differences (Bower and Haley, 1964; Sampson, 1969). In view of the findings reported here it is possible that the differences appeared largely as a result of protracted performance. Certainly the differences between the nasal and temporal hemiretinae increased considerably towards the end of the test period. In other words the temporal hemiretinae appeared to maintain a high level of performance in the vigilance situation while nasal performance declined. One factor in the maintenance of sensitivity in the temporal, but declining sensitivity in the nasal, hemiretinae seems related to the maintenance of central vision and the deterioration of peripheral vision on tasks of this type (Hockey, 1969). It need hardly be pointed out that the progressive decline of sensitivity of the nasal hemiretinae provides a simple explanation of the tunnel vision effect as an alternative to supposing that it arises as a consequence of the pre-emptive use of central vision to observe the most important events.

It is somewhat surprising that the nasal hemiretinae should be the ones to show the decline in sensitivity. It is often assumed that in the optic system the crossed fibres are more effective than the uncrossed (Bower, 1966; Bower and Haley, 1964). Bower and Haley showed that when stimulus durations are very brief in conditions involving binocular rivalry, the rivalry is resolved in favour of the nasal hemiretinae. Sampson (1969) reported that recall of digits projected to nasal hemiretinae was significantly greater than when projected to temporal hemiretinae. However, Crovitz and Lipscomb (1963) also reported that the temporal visual fields show a greater sustained performance than the nasal during short durations of stimulation. The results described here suggest that while the capacity and sensitivity of the nasal hemiretinae may at first be the equal of or even exceed that of the temporal, when more sustained performance is undertaken, the nasal is definitely inferior. There is no reason to suppose that one hemiretina performing initially at a high level must necessarily also have the greater capacity for continuing performance, and that this is not so was clearly demonstrated by the present experiment.

The work reported here was undertaken during tenure of grant G969/96C of the Medical Research Council. The authors are greatly indebted to Philip Kinsey for his assistance both in the construction of apparatus and in the conduct of the experiment.

References

Bower, T. G. R. (1966). A local sign for depth. *Nature*, **210**, 1081–2.

Bower, T. G. R. and Haley, L. J. (1964). Temporal factors in binocular vision. *Psychonomic Science*, **1**, 409–10.

Crovitz, H. F. and Lipscomb, D. B. (1963). Dominance of the temporal visual fields at short durations. *American Journal of Psychology*, **76**, 631–7.

Dimond, S. J. (1970). Hemispheric refractoriness and the control of reaction time. *Quarterly Journal of Experimental Psychology*, **224**, 610–7.

Dimond, S. J. (1971). Hemispheric function and word registration. *Journal of Experimental Psychology*, **87**, 183–6.

Dimond, S. J. and Beaumont, J. G. (1971). The use of two cerebral hemispheres to increase brain capacity. *Nature*, **232**, 270–1.

Hockey, G. R. J. (1970). Changes in attention allocation in a multicomponent task under loss of sleep. *British Journal of Psychology*, **61**, 473–80.

Levy-Agresti, J. and Sperry, R. W. (1968). Differential perceptual capacities in major and minor hemispheres. *Proceedings of the United States National Academy of Science*, **61**, 1151.

McGrath, J. J. (1963). Some problems of definition and criteria in the study of vigilance performance. In Buckner, D. N. and McGrath, J. J. (Eds.), *Vigilance, A Symposium*. New York: McGraw-Hill.

Sampson, H. (1969). Recall of digits projected to temporal and nasal hemi-retinas. *Quarterly Journal of Experimental Psychology*, **21**, 39–42.

White, M. J. (1969). Laterality differences in perception, A review. *Psychological Bulletin*, **72**, 387–405.

Received 1 July 1971

Quarterly Journal of Experimental Psychology **23**, 449–452

THE EFFECT OF SLEEP DEPRIVATION ON SIGNAL DETECTION PARAMETERS

MARYANN DEATON, J. S. TOBIAS† AND R. T. WILKINSON‡

Psychological Laboratory and the Applied Psychology Unit, Cambridge

Twelve subjects performed a high signal rate vigilance task, once after a night's sleep, and once after a night without sleep. Raw scores were transformed into the signal detection parameters, d' and β. After sleep deprivation, detection performance was significantly impaired. This was reflected in a fall of d', whereas β was not significantly altered. Analysing the control data alone for comparison with other vigilance studies revealed a decrement in % signals detected and d', and an increase in β from the first to the second half of the test.

Introduction

Much recent work has shown that performance on various tasks is impaired after sleep deprivation (Wilkinson, 1958; Williams, Lubin and Goodnow, 1959; Wilkinson, 1961; Williams, Kearney and Lubin, 1965). However there has been little conclusive evidence as to *why* this happens. Williams *et al.* (1959) suggest that it is due to momentary lapses in attention to the task, the number of these lapses increasing after sleep deprivation. Clearly this idea is compatible with Broadbent's (1958) view that a selective filter is operating at the entrance to a channel of limited capacity, and that shifts of the filter to irrelevant stimuli could account for such lapses. Further, Bjerner (1949) noticed that lapses during a task were accompanied by alpha wave depression; it has therefore been suggested that such lapses represent "microsleeps".

A different way of approaching the problem involves the use of signal detection theory (Swets, Tanner and Birdsall, 1961). The advantage of this theory is that it provides a means of disassociating impairment due to an intrinsic loss of sensitivity (d') from that due to increased caution in deciding that a given event is a signal one (β), (Howland, 1958; Broadbent and Gregory, 1963; Mackworth and Taylor, 1963). A present theory of sleep deprivation (Murray, 1965) suggests that it increases the drive to sleep at the expense of all other drives. It would seem unlikely that such a change should alter an organism's ability to make basic sensory discriminations, an ability which d' is taken to reflect. On this basis, then, d' should remain invariant with loss of sleep, leaving β, perhaps, to reflect the changes in organization responsible for any decline in performance. Wilkinson, Edwards and Haines (1966) and Wilkinson (1968) have already reported findings contrary to this hypothesis, namely a fall in d' and no change in β with one night of sleep deprivation. The vigilance task in this setting required detection of signals which

† Now at St. Bartholomew's Hospital, London.
‡ Senior author.

formed a very low proportion of the stimuli presented, the signal to non-signal ratio being 0·022. This is very different to the original setting of Swets *et al.* where probabilities of signal + noise and noise only presentations are equal. One of the aims of the present experiment has been to see whether the fall in d' with one night of sleep deprivation persists when signal probability is increased to make the task more like the original setting of Swets *et al.*, while still retaining the format of a vigilance task.

Method

Subjects and apparatus

The subjects were 12 enlisted men whose ages ranged from 19 to 25 and who had little previous experience of total sleep deprivation. They carried out a vigilance task three times, once as a practice session, once after normal sleep, and once following a night of no sleep (33 hr sleep deprivation). The order of the tests with and without sleep was balanced across subjects. Each test lasted 30 min, during which subjects heard $\frac{1}{2}$ sec tones, well above threshold, at 2 sec intervals, against a background of moderate white noise. The task was to detect tones which were slightly shorter than the rest. These "signals" occurred at random intervals, but very frequently by vigilance standards, i.e. on average, one every four tones. Subjects reported these signals by pressing a key as quickly as possible. They were given knowledge of results (signals detected and false reports) following each test.

Results

For each subject a record was taken of correct signal detections and false reports in the first and second half of the test under conditions of normal sleep and sleep deprivation. From these scores d' and β were derived for such subject using Freeman's (1964) Tables.

Only two aspects† of the results are considered, (1) the effect of sleep deprivation over the whole test and (2) deterioration in performance from first to second half of the test under conditions of normal sleep only (to allow comparison with other studies of vigilance decrement under normal conditions). Specifically, the question here was whether a vigilance decrement occurred under these high signal frequency conditions, and if so, whether it was reflected in changes of d', or β.

The proportions (normal ÷ sleep deprived score) and (first half of the test ÷ second half) were assessed for significance by the Wilcoxon Test (Siegel, 1956) the proportions being first transformed into $\log_{10}$ to achieve symmetrical distribution about zero. The significant results were:

(1) a decrease in the percentage of correct detections with sleep deprivation ($P < 0{\cdot}01$) (Table I);

(2) a decrease in d' from the normal condition to the sleep deprived ($P < 0{\cdot}01$) (Table I);

(3) a decrease in d' from the first to the second half of the control test ($P < 0{\cdot}05$) (Table II);

(4) an increase in β from first to second half of the control test ($P < 0{\cdot}02$).

† This was an undergraduate project and time did not permit further analysis, e.g. vigilance decrement as a function of sleep deprivation. However, this has been thoroughly studied in the literature (see Wilkinson, 1965).

TABLE I

Effect of sleep deprivation on detection, false reports, d′ and β, averaged over 12 subjects

	Probability of a correct detection	Probability of a false report	d'	β
Normal	0·67	0·046	2·38	7·18
Sleep Deprived	0·54	0·051	1·85	4·19

TABLE II

Vigilance decrement. Comparison of scores (as in Table I) for first and second halves of test averaged over 12 subjects in normal tests only

	Probability of a correct detection	Probability of a false report	d'	β
Normal first half	0·73	0·053	2·50	6·48
Normal second half	0·62	0·039	2·33	9·49

Discussion

It is clear from our results that the major effect of sleep deprivation in this situation is to decrease d'. This suggests that intrinsic capability is impaired. Wilkinson (1968) found d' reduced by sleep deprivation, using a vigilance test with a relatively low signal frequency, which lasted an hour, and which was administered five times during the day following sleep deprivation. The present test involved a hectic, high rate of signal frequency, lasted 30 min, and was administered only once. That it, too, yielded a fall in d' with sleep deprivation confirms the earlier result in a TSD setting more akin to the original of Swets, *et al.* (1961), in that the signal and no-signal event probabilities were closer than is usual in vigilance situations.

The results in the control condition of this experiment may be considered separately in relation to other evidence on the decline of performance on a vigilance task with prolonged work, and in particular with regard to the question whether this decline is associated with a fall in d' or a rise in β. In the present experiment both occurred. The relevant parameters were an event rate of 30 per min, a signal rate of 7·5/min, a signal/event ratio of 0·25, and auditory mode of presentation.

A similar decline in d' occurred in an experiment by Loeb and Binford (1968) where one of their conditions had certain parameters similar to ours; viz. event rate 24/min and auditory presentation. On the other hand their signal rate (0·5/min) and signal-to-nonsignal ratio (0·021) were very different. This may give some reason for thinking that event rate and modality are the important factors in deciding the extent to which d' declines during a vigilance test. That modality

may be the more important of the two is suggested by the failure of both Broadbent and Gregory (1963) and Mackworth (1965) to obtain a similar decline of d' tasks where the event rate was about the same as ours but the modality a visual one. The present results are also in line with current thought on coupling in vigilance tasks (e.g. Hatfield and Loeb, 1968) to the effect that the closely coupled task, i.e. where subjects cannot easily avoid being stimulated by the task material, is the one most likely to show a decline in d' during its course. Our task, being an auditory one, was closely coupled by this definition and conformed to this pattern.

References

BJERNER, B. (1949). Alpha depression and lowered pulse rate during delayed actions in a serial reaction test. *Acta Physiologica Scandinavica* **19**, *Suppl.* No. 65.

BROADBENT, D. E. (1958). *Perception and Communication.* London: Pergamon Press.

BROADBENT, D. E. and GREGORY, M. (1963). Vigilance considered as statistical decision. *British Journal of Psychology* **54**, 309–23.

FREEMAN, P. R. (1964). *Table of d' and β.* Medical Research Council, Applied Psychology Unit, Report No. A.P.U. 529/64.

HATFIELD, J. L. and LOEB, M. (1968). Sense mode and coupling in a vigilance task. *Perception and Psychophysics*, **4**, 29–36.

HOWLAND, D. (1958). An investigation of the performance of the human monitor. USAF WADC Technical Report No. 57–431.

LOEB, M. and BINFORD, J. R. (1968). Variation in performance on auditory visual monitoring tasks as a function of signal and stimulus frequencies. *Perception and Psychophysics*, **4**, 361–4.

MACKWORTH, J. F. and TAYLOR, M. M. (1963). The d' measure of signal detectability in vigilance-like situations. *Canadian Journal of Psychology*, **17**, 302–25.

MACKWORTH, J. F. (1965). Deterioration of signal detectability during a vigilance task as a function of event rate. *Psychonomic Science*, **3**, 421–2.

MURRAY, E. J. (1965). *Sleep, Dreams and Arousal.* New York: McGraw-Hill.

SIEGEL, S. (1956). *Non-parametric Statistics.* New York: McGraw-Hill.

SWETS, J. S., TANNER, W. P. and BIRDSALL, T. G. (1961). Decision processes in perception. *Psychological Review*, **68**, 301–40.

WILKINSON, R. T. (1958). The effects of sleep loss on performance. Medical Research Council, Applied Psychology Unit, Report No. 323/58.

WILKINSON, R. T. (1961). Interaction of lack of sleep with knowledge of results, repeated testing and individual differences. *Journal of Experimental Psychology*, **62**, 263–71.

WILKINSON, R. T., EDWARDS, R. S. and HAINES, E. (1966). Performance following a night of reduced sleep. *Psychonomic Science*, **5**, 471–2.

WILKINSON, R. T. (1968). Sleep deprivation: Performance tests for partial and selective sleep deprivation. In: Abt, L.A. and Riess, B.F. (Eds), *Progress in Clinical Psychology*. Vol. 7. New York: Grune and Stratton.

WILLIAMS, H. L., LUBIN, A. and GOODNOW, J. J. (1959). Impaired performance with acute sleep loss. *Psychological Monographs*, **73**, No. 14.

WILLIAMS, H. L., KEARNEY, O. F. and LUBIN, A. L. (1965). Signal uncertainty and sleep loss. *Journal of Experimental Psychology*, **69**, 401–7.

Received 6 *August* 1971

Quarterly Journal of Experimental Psychology (1971) **23**, 453–461

BOOK REVIEWS

Gaze, R. M. *The Formation of Nerve Connections.* London and New York. Academic Press. 1970. Pp. 282. £4.00.

This important monograph studies the way in which the processes of nerve cells, proliferating in the developing animal, can so direct their growth and recognize their destination as to make correct functional contact with their targets, often in quite remote parts of the body. It is in a field that has produced imaginative theories and skilled experiments for nearly 50 years, and seems now to be assuming greater importance in the explanation of more general aspects of neural function.

It is unlikely that tissues other than the nervous system develop according to very different principles, so to this extent the book is concerned with embryology in general. To some, the topological organization of the central sensory representations, and the guidance of developing or regenerating axons over long distances have seemed ideal paradigms for experimental embryology. But though the nervous system seems to provide some tidy developmental problems, it has not yet generated basic ideas beyond the well-used classical ones of gradient and polarity. Perhaps it only seemed appropriate in the first place in the light of these concepts.

Gaze does not dwell for long on the wider theoretical aspects of development, but lays out a close interpretation of the frequently confusing experimental analysis of neural development and regeneration. If his book has a theme, it is to emphasize the structural intricacy and specificity of nervous tissue, and to defend the position that its function is substantially embodied in its detailed anatomy.

In the first two chapters, he reviews experiments on the development of nerve fibre connections between the peripheral sensory and motor apparatus and the centre. This is largely an historical account, shaped to examine the classical modulation theory of Paul Weiss in the light of many years of experimental evaluation.

Weiss's notion, that the central connections of nerve cells can be determined by the identity of the peripheral organs they come to innervate, is economic and compelling in that it removes the need to suppose that each developing nerve cell is provided with explicit instructions on the form and connectivity that it is to assume. However, with one or two inadequately documented exceptions, Gaze concludes that most of the classical experiments are open to other interpretations, and urgently need detailed anatomic study. In a way this is disappointing, for the major assertion of Weiss's theory is no less plausible than that the cells of the head are determined to a degree by the anterior pole environment into which they first cleave. However, this conservative and sceptical account indicates the level of detail at which transplant and redirection experiments need to be interpreted.

In the second half of the book, Gaze deals with the area in which he and his co-workers have been active over the last 12 years. The third chapter is concerned with the fibre projection from the retina to the tectum in amphibia, and the final one with the connections between the two halves of the tectum that subserve binocular integration.

R. W. Sperry first interfered surgically with the eyeball and optic nerve in amphibians, and observed the permanent maladaptive consequences that accrued for the animals' vision. He was led to the powerful generalization that optic nerve fibres, as they grow into the tectum in development or regeneration, seek out appropriate target loci matching the unique and inflexible "place" labels that each one carries. It was an idea largely incompatible with that of Weiss, but relevant to a quite different system.

Gaze, with Jacobson and Szekely, modified this conclusion in two respects. Evidently the labels are acquired by the retinal neurones at a critical stage of development, after the eye rudiment is well established, and they are specified separately (at different times) with respect to the horizontal and vertical co-ordinates of the eye. It is as if they are read off

the co-ordinate values of a pair of crossed quantitative gradients. This sort of mechanism has the same attractive parsimony as Weiss's theory, for a large number of labels are transcribed not directly from the genetic material, but indirectly from the co-ordinate system whose own specification would be relative simple.

Just as behaviour is held to be the outcome of structured connectivity and excitation flow between cells of the nervous system, so it is thought, with surprisingly little evidence beyond common sense, that its adaptive modification reflects alteration in this functional anatomy. Gaze describes experiments that exploit the immutably specified connections between amphibian eye and tectum, and then goes on to consider the quite recently discovered plasticity in the neural projection between the two halves of the tectum.

His group have recently shown that the inter-tectal neurones, which link together corresponding points on the representations of each eye field, develop and maintain their connections in a way suggesting that they do not grow to matching place labelled neurones, but rather to loci where visually evoked activity, originating in one eye, matches exactly their own, generated in the other. It has been known for some time that development of the mechanisms of stereopsis is critically susceptible to visual experience. This work shows further that experiential factors as subtle as the conjunction of activity in growing fibres and their targets seem able to guide large scale arrangement of connections between the neurones involved.

These beautiful findings are important at two levels. Minimally they show another way in which the extraordinary structural richness of the nervous system is elaborated in development. Further, perhaps, if this contiguity guidance of growth is not merely a property of neurones associated with the two eyes, they might provide a prototype for the plasticity of other nervous functions.

The pleasure of reading this book is not only due to the importance of the subject matter. It is as much a progress report as a definitive account, written lucidly with emphasis on what is uncertain at the expense of what is established. Molecules are hardly ever mentioned, and it has no need to borrow jargon from other disciplines. It is even humorous in places. It should certainly be an inspiration in its field.

JOHN SCHOLES

CORNSWEET, T. N. *Visual Perception.* New York and London: Academic Press. 1970. Pp. xi + 475. £7.00.

Although one would probably not think so from the title, this book is an introduction to human visual psychophysics. It deals with many of the usual topics in this field, such as quantal effects at absolute threshold, increment thresholds, colour vision, contrast modulation transfer functions, and flicker fusion. A little visual physiology is also included, which, as might be expected in a book of this type, is largely confined to work on *Limulus*.

There are, of course, a large number of other texts that deal with these same topics. The present book is exceptional, however, for the clarity with which they are explained. In order to achieve this the book starts by working step by step through a description of Hecht, Schlaer, and Pirenne's classic experiment on the absolute threshold. This description occupies almost 80 pages, and is used to introduce many basic visual concepts, such as depth of focus, image formation, the structure of the eye, and spectral sensitivity. On occasion the detailed treatment of very simple points is rather tedious, but the resulting explanations should be readily comprehensible to most undergraduates, which is certainly not true for the majority of books on vision.

The later sections of the book proceed rather more rapidly, although each point as it arises is still treated in great detail. The section on contrast modulation transfer functions forms a particularly clear and useful introduction to this topic, especially as most of the research in this area is relatively recent and not described in the standard texts. The chapters on colour vision are also unusually good. The approach in this case is unconventional and both far easier to understand and much more interesting than most descriptions of this topic. It is a pity, however, that the author chose to relegate the C.I.E. system of colour

measurement to a footnote, since it would have been relatively easy to provide an adequate description of this following the very lucid description of colour that he does give. Irrespective of whether one feels that the C.I.E. system is unnecessarily obscure or unusually elegant and satisfying, the fact remains that it is widely used and that much of the literature in the field of colour cannot be understood without it.

In summary, therefore, the book gives a valuable and exceptionally clear introduction to some fields in vision, and any undergraduate would be well repaid for reading it. It is aimed at, and useful to students, not research workers, and it is a pity that the price is quite beyond the purses of the public for which it was written.

W. R. A. MUNTZ

ESTES, W. K. *Learning Theory and Mental Development*. New York and London: Academic Press. 1970. Pp. 223.

"The word 'and' in the title of this volume signifies a logical product, not a logical sum." Thus Estes in the opening sentence of this book. As he argues, to survey either modern learning theory or the development of intelligence alone (in fact the book is more specifically concerned with causes of mental retardation) would require a substantial volume, "but the interaction of learning theory with the study of mental development has been so slight that an extensive search of the literature yields material for only a slender volume."

The remainder of the book amply documents the truth of this last assertion. Indeed, had Estes felt rigorously bound by his initial restriction, the volume would have shrunk to the dimensions of a substantial journal article. As it is, large sections of the book are in fact devoted to (a) an informal presentation of the type of theorizing about learning that Estes has recently espoused, as well as of other theories such as Hull's; (b) a chapter devoted to a review of "studies demonstrating variables or procedures which produce large effects upon learning rates"—which turns out to be a somewhat haphazard collection of recent studies which have captured Estes' interest; and (c) a 36-page chapter devoted to an exposition and analysis of Zeaman and House's model for discrimination learning, only five pages of which discuss its application to the behaviour of retarded subjects.

It is no fault of Estes' that the ostensible subject matter of his book should be so meagre. It remains a pity that what he does have to say on the subject should be interspersed among matter that is either irrelevant or more satisfactorily treated elsewhere. For as one would expect from its author's stature, this book contains a number of acute and penetrating judgements, both on the methodological problems involved in comparing retarded and normal populations, and on the adequacy of the theoretical analyses that have been offered. Commenting on the former, he is particularly insistent on "the conclusion that when differences in previous opportunities to learn and the relationship of these to the reference task are unknown, comparisons between different groups of individuals are of little value for any theoretical purpose." And the researcher who points to the difficulties inherent in providing appropriate controls is asked to remember "the fact that better controls will be difficult to achieve adds nothing to the value of uninterpretable data". In the end, Estes is able to give even qualified approval to only two ideas intended to explain the deficiencies from which retarded subjects suffer: they have poor rehearsal strategies (Ellis) and do not attend to the relevant cues of visual discrimination problems (Zeaman and House). It is a sad commentary on much of modern learning theory that it has had no place for these sorts of concepts, but this fact doubtless serves to explain why the material for this book is so slight.

N. J. MACKINTOSH

SCHOENFELD, W. N. (Ed.). *The Theory of Reinforcement Schedules*. New York: Appleton-Century-Crofts. 1970. Pp. x + 316. (Price not known.)

This book results from a symposium on schedules of reinforcement held in 1966. As is usual with such publications when no strong editorial policy has been followed, the quality of the individual contributions varies greatly, as does their importance.

An observer of operant conditioning might characterize the field as an apparently endless description of differences in patterns of behaviour maintained by schedules which differ only trivially in the rules by which they deliver reinforcements. Precise description of such rules has long proved cumbersome in normal language, and so jargon proliferates. There have been suggestions for more formal or symbolic representations of these rules (by Skinner and by Mechner, for example), and Snapper *et al.* (Ch. 8) here offer a further system based on state graphs. This will probably achieve no wider a usage than the others.

The traditional point of view concerning the effects of intermittent schedules of reinforcement is that they result from the interaction of a number of simple processes which have been documented separately. Among these are the strengthening effects on behaviour of the delivery of a reinforcer, extinction effects during periods of non-reinforcement, and the development of discriminative control by differential reinforcement in the presence of certain stimuli. However, such conditions as the latter are frequently not programmed by the experimenter, who may therefore only infer them in order to explain schedule-controlled behaviour. Two chapters in this book appear to be within this general tradition. Dews (Ch. 2) elaborates the interpretation of fixed-interval behaviour as being due to differential delay of reinforcement which many will already associate with him. Morse and Kelleher (Ch. 5) present their view that schedules *per se* are fundamental determinants of behaviour. They suggest that it is the schedule which determines the patterning of operant behaviour and that it is this patterning (in terms of local rates of responding) which determines the effects of additional independent variables, such as drugs. However, their case is weakened at present, in this reviewer's opinion, by the difficulty of isolating the crucial variables in some of the experiments cited to support this view.

In a particularly important discussion, Jenkins (Ch. 3), on the other hand, argues that schedules are merely contrivances which may be useful in some circumstances: "neither man nor animals are found in nature responding repeatedly in an unchanging environment for occasional reinforcement" (p. 107). Such a view depends to some extent on one's interpretation of a unit of behaviour (discussed more explicitly by Schoenfeld and Farmer in Ch. 7); however, Jenkins' insistence on the importance of *sequences* of environmental events and their relationship to behaviour is clearly of considerable conceptual importance. Greater attention to such sequential effects is surely required if a satisfactory "theory" of schedules (both of primary and of conditioned reinforcement) is to evolve, although one might not agree with Jenkins' view that such sequential effects are best elucidated by discrete trial procedures. However, one might also argue that Weiss' advocacy (Ch. 9) of what he calls a microanalysis of *transitional* states of behaviour is not the solution, since this may fail to come to grips with crucial causal determinants of behaviour. In general, discrete trials procedures may be more appropriate to elucidate variables in some situations, but free-operant schedules may be more useful in other contexts, a point made by Logan and Ferraro (Ch. 4), who argue that the two procedures are distinguishable, but complementary.

The use of discrete trials to analyse schedule effects is strongly supported, in effect, by Ray and Sidman (Ch. 6). They agree with others that schedules may favour the development of behavioural control by stimuli which cannot be directly manipulated by the experimenter. Thus, the effects of exteroceptive stimuli may be contaminated by control deriving from different sources, such as the preceding behaviour. In this way, these authors are also led to an analysis of schedule effects by means of discrete trial procedures. This conclusion is of particular interest when advocated by one whose earlier work on schedules of free-operant avoidance may be regarded as seminal in the development of the radical theories of aversive control currently being explored (by Herrnstein, for instance). Catania's consideration (Ch. 1) of temporal properties of behaviour is yet another in which a procedure analogous to a discrete trials analysis is to be found discussed.

This book, then, contains data and discussions of considerable interest to operant conditioners, some of which is extremely challenging to what has become almost dogma in this area of research. Certainly, many of the opinions expressed reflect dissatisfaction, although the suggested remedies may themselves be marred by difficulties. The Editor describes this book as a "remarkable diversity of interests, approaches, treatments and definitions"

(Preface, p. vi). But the style of presentation of many of the chapters, and the organization of the book as a whole, provides little help to a reader attempting to achieve some degree of synthesis. Moreover, the breadth of coverage is less than satisfactory (as the Editor himself notes) and some of the chapters already seem somewhat dated in the light of more recent findings. The book is therefore not to be strongly recommended to the non-specialist—the "theory" suggested in its title may yet be a long time coming.

DEREK BLACKMAN

REESE, HAYNE W. AND LIPSITT, LEWIS P. *Experimental Child Psychology*. New York and London: Academic Press. 1970. Pp. 782. £6.30.

We are told in the preface that "this textbook was written for courses in experimental child psychology at the advanced undergraduate and early graduate level". Upon completing this book, however, one is hard put to find any one level at which the book might be of use. The overriding impression is one of unevenness. This is no doubt due to the fact that Reese and Lipsitt are not the sole authors; rather they are editors who have collated and re-written, changed and added to, chapters contributed by 16 other authors. The interested reader is forced to consult the preface in order to discover the author of each particular chapter.

It is probably this partly hidden authorship that leads to the extreme variation in the quality of this book. Thus, some chapters like 15 and 16 on intelligence, by W. O. Shepard, are far too introductory and do not deal with enough specifically experimental studies. The chapter on cognitive development (chapter 14) by D. Elkind is surprisingly poor; it is limited to a discussion of aspects of Piaget's work that any modern *general* introductory text already adequately covers. Other chapters, however, seem not only to be more advanced but are excellent for demonstrating experimental methodology. Chapters 3 and 4 on the basic learning processes, written mainly by E. R. Siqueland, are excellent from this point of view; the very good analysis of experiments as being well or badly performed and adequately or inadequately controlled is, however, somewhat spoiled by the failure to convey a sense of unity to the material. At an opposite pole, there are chapters like 17 on socialization, by McCandless, and 18 on biogenetic factors in development, by Glickstein and Moore, which do a very good job of introducing and discussing a topic area but which are poor in their mention and evaluation of experimental studies. The chapter on language acquisition by Palermo (chapter 13) and most of that on perception by Wohlwill (chapter 11) manage to combine a clear presentation with excellent information from properly evaluated experiments. Chapter 10 on motivation, written mainly by Longstreth, seems also to strike a nice balance between presentation of a topic area and the experimental work supporting it. His technique has been to choose a few illustrative experiments and to deal with these at length, explaining the concepts used as he goes along.

It appears that in an effort towards comprehensiveness, several chapters were "tacked on". By far the worst of these is Lipsitt's contribution on emotional development (chapter 12). Experimental psychologists will only have their prejudices against this field reinforced by this most perfunctory treatment of the topic. Very few experiments are mentioned, and many of these are 40 years old. Entire topics such as aggression, frustration, anger, anxiety and fear, are left unmentioned. There is no citation of any of the experimental literature in childhood emotional disorders. Indeed, this chapter reads like a hurriedly written response to the suggestion that "emotional development" must not be entirely ignored by a text covering all of experimental child psychology. And of course, an introduction must be written. This introductory chapter about the scope and methods of experimental child psychology is a perfect example of what a reviewer in these pages (November, 1970) wrote about such chapters in general, which can be paraphrased: Most chapters on experimental design are ponderous and dull; they do not form a basis for what follows, nor does formalizing and classifying research tactics promote good research.

A major fault of the book may lie in the choice of the experiments which are reported. This is a difficult problem. When a person involved in research reviews an experimental area of psychology, it is not altogether surprising that he will choose to emphasize those

September 1969. It is divided into three sections. Section (1) on **man** describes the psychological and physiological effects of stress at work. Section (2) on **techniques** covers the measurement of performance, the use of additional tasks, and subjective assessments by workers and investigators. There are also electrophysiological techniques, and the inevitable Douglas bags and bicycle ergometers. Section (3) on **applications** describes the personal experiences of a number of eminent ergonomists.

Each section starts with a prologue and ends with an epilogue, in which attempts are made to link the individual contributions together. There is an excellent introductory paper by Chapanis, which collects together a number of his recent poignant comments about laboratory experiments on ergonomics. An attempt is made at the end to pull together many of the divergent points made at the symposium.

Proceedings of symposia have the advantage of being more up to date than collections of readings, which are likely to have been published for the first time in journals anything up to 10 or 15 years previously. But there are disadvantages. The organizers select the contributors, not the contributions. A number of individual contributions cover the same ground. Also one has a feeling that parts of the book would have appeared under one guise or another, whatever the title of the symposium. This increases the difficulty of the editors in attempting to pull the contributions together.

In this particular collection, many of the contributions were clearly written directly for the symposium. The book is a useful introduction for students of ergonomics. It is of interest also to the general reader, provided he is concerned with measurement and practical applications, rather than with finding out why things are the way they are.

The main criticism is the unergonomic printing and layout of the book. The lines of print are almost 6 inches long. With the 10-point type used, this means 3½ alphabets to the line. It is as bad as Burt's original layout of the *British Journal of Statistical Psychology*. The headings and subject index are printed mainly in italics, which Tinker and Paterson found to be illegible as far back as 1928. The first page of each prologue is printed on dark brown paper, which has a subjective reflectance in dim light of about 8%. It is barely legible. Surely the printing of a book on ergonomics should follow the ergonomic principles which it expounds!

E. C. POULTON

NELSON, D. *SSRC: Bibliography of British Psychological Research 1960-1966*. Her Majesty's Stationery Office. 1971. Pp. 210. £1.00.

This bibliography contains a reasonably complete list of publications by British psychologists in the years 1960–66; it is marred by numerous inaccuracies. The references are set out under some 45 different topics, and each section contains a brief introduction written by 45 unnamed "experts" who assess British work in the area. The assessments make depressing reading. "The overall reaction to the picture presented by British work is one of some gloom" (p. 65), "the standard needs to be set much higher in this field" (p. 108), "our research contribution is microscopic" (p. 112). Not everyone will agree that "the contribution of British psychologists to theoretical psychology is exceptionally weak, and has no marked characteristics except perhaps moderation, or even mediocrity" (p. 1) and we may perhaps be thankful rather than sorry that "there is no enterprise comparable to the study of psychology directed by Koch in America" (p. 1).

The great weakness of British experimental psychology would appear to be a lack of professionalism rather than any lack of ideas. American investigators tend to devote more attention to proper experimental design, the use of correct statistics, careful control of possible contaminating factors, and to the careful preparation of articles for publication. In all these respects British work is often too slipshod and lacking in care though the ideas behind the work may often be highly original. It is hard to say whether the reason for this defect lies in our inadequate training of graduates or whether it is merely part and parcel of the British love of the amateur.

No raison d'être is given for the production of this volume and one wonders whether SSRC would not further the subject more by spending money to support research rather than by sponsoring this sort of enterprise.

N. S. Sutherland

Publications Received

Bailey, D. E. *Probability and Statistics: Models for Research.* Chichester: John Wiley. 1971. Pp. 686. £5.75.

Connolly, K. J. (Ed.). *Mechanisms of Motor Skill Development.* London: Academic Press. 1971. Pp. 393. £6.00.

Demling, L. and Ottenjann, R. (Eds.). *Gastrointestinal Motility:* International Symposion on Motility of the GI-Tract, Erlangen, July 15th and 16th, 1969. Stuttgart: Georg Thieme Verlag. 1971. Pp. 219. DM 27,50.

Ewen, R. B. *Introductory Statistics for the Behavioral Sciences Workbook.* New York: Academic Press. 1971. Pp. 155. £1.40.

Ford, D. H. and Schade, J. P. *Atlas of the Human Brain.* Second Revised Edition. London: Elsevier. 1966. Pp. 233. £4.25.

Goldschmid, M. L. (Ed.). *Black Americans and White Racism.* New York: Holt, Rinehart and Winston. 1970. Pp. 434. £2.90.

Herriot, P. *Language and Teaching: a Psychological View.* London: Methuen. 1971. Pp. 109. £1.25.

Johnsen, E. G. and Corliss, W. R. *Human Factors Applications in Teleoperator Design and Operation.* Chichester: John Wiley. 1971. Pp. 252. £6.25.

Kaltenbach, M. and Lichtlen, P. (Eds.). *Coronary Heart Disease: International Symposium in Frankfurt, January* 22-24, 1970. Stuttgart: Georg Thieme Verlag. 1971. Pp. 271. DM 39,50.

Loewenstein, W. R. (Ed.). *Handbook of Sensory Physiology. Vol. 1.* Berlin: Springer-Verlag. 1971. Pp. 600. $46.20.

Luthe, W. *Autogenic therapy, Vol. V.* New York, London: Grune & Stratton. 1970. Pp. 344. $17.50.

McFarland, H. S. N. *Psychological Theory and Educational Practice.* London: Routledge and Kegan Paul. 1971. Pp. 323. £2.25.

Miller, N. E. *Selected Papers.* Chicago: Aldine Publishing. 1971. Pp. 874. £15.00.

Morton, J. (Ed.). *Biological and Social factors in Psycholinguistics.* London: Logos Press. 1971. Pp. 215. £3.50.

Piaget, J. *Structuralism.* London: Routledge & Kegan Paul. 1971. Pp. 153. £2.00.

Vernon, M. D. *Perception through Experience.* London: Methuen, New York: Barnes and Noble. 1970. Pp. 306. £2.75.

Welkowitz, J., Ewen, R. B. and Cohen, J. *Introductory Statistics for the Behavioral Sciences.* New York: Academic Press. 1971. Pp. 271. £3.95.

CONTENTS

Printed by W Heffer & Sons Ltd Cambridge England